Norwegian University of Life Sciences

Master's Thesis 2022 60 ECTS

Faculty of Chemistry, Biotechnology and Food Science
Daniel Straume

Characterization of a novel cell wall binding domain of the fratricin CbpD from *Streptococcus dysgalactiae*

Jennie Ann Allred
Biotechnology

Acknowledgments

This thesis was completed as part of the Master's programme in Biotechnology at the Norwegian University of Life Sciences (NMBU). The work was conducted in the Molecular Microbiology research group at the Faculty of Chemistry, Biotechnology and Food Science (KBM) between August 2021 and May 2022.

I extend tremendous appreciation to my main supervisor Dr. Daniel Straume, for his excellent guidance throughout the project. This includes advice and assistance in the laboratory and answering all thesis-related questions. As a result, I have learned a lot, which surely will be of great use in the future. I would also like to thank my co-supervisor Dr. Morten Kjos, for always being around to help, and Zhian Salehian for having the patience to repeatedly demonstrate the various techniques and machines in the lab (especially SDS-PAGE). Overall, the research group consists of kind and helpful individuals who have created a great work environment and been highly supportive during the year.

I would also like to express gratitude to my fellow master students (Rebekka, Tiril, Ingrid, Anne and Vilde), where many hours have gone to talking, but also helping each other out when needed. The morale has been high among us, partially due to fun social events and positive attitudes throughout.

Lastly, I would like to thank my family, my dad and brother for looking over my thesis, my mom for unconditional support and my twin sister for letting me call whenever. A special thanks to my friends and partner, who have been kind and stuck by me through thick and thin.

Jennie Ann Allred,
Ås, May 2022

Abstract

The *Streptococcus* genus contains a large variety of species across seven subgroups. The lesser-known *Streptococcus dysgalactiae* of the pyogenic group is both a human and an animal pathogen and is subsequently divided into two subspecies: *S. dysgalactiae* subsp. *equisimilis* (SDSE) and *S. dysgalactiae* subsp. *dysgalactiae* (SDSD), respectively. The SDSD has been the main focus of the present work, where it has great significance in causing bovine mastitis in cows and septic arthritis in lambs. This has caused various problems in the dairy industry and livestock production, where antibiotics for treatment are necessary. The relatively recent observation of zoonotic diseases in humans, and reports of penicillin-resistant SDSE isolates, have made it a vital pathogen to focus our studies on. In addition, considering *S. dysgalactiae* has a complete set of competence genes and a fratricin for competence-induced cell lysis, natural competence for genetic transformation might be an important horizontal gene transfer mechanism to acquire genes for adaptability and evolution. The following work has therefore been an in-depth investigation of the *S. dysgalactiae* fratricin protein and specifically on its lesser characterized conserved domain. It has previously been shown by fluorescence microscopy that the protein binds to the septal region (the division zone) of targeted cells. Here, it was seen that by fusing a superfolder green fluorescent protein onto the conserved domain, it still retained its septal binding abilities, demonstrating that this domain is responsible for directing the fratricin to septum. A conserved leucine-alanine-glycine-glycine motif and conserved amino acids glutamine and tryptophan were identified by a multiple sequence alignment, individually point-mutated by overlap-extension polymerase chain reaction and visualized by a fluorescence microscope to determine if the septal binding patterns changed. Unevenly binding across the entire target cells was observed, indicating that these highly conserved amino acids are important for the conserved domain's ability for septal recognition. Several protocols for removal of teichoic acids on isolated *S. dysgalactiae* sacculi were also performed, where using trichloroacetic acid for two weeks on the sacculi was ultimately chosen. Sacculi deficient in teichoic acids showed no septal recognition by the fratricin, revealing that specific molecular moieties on these polymers are necessary to guide the protein towards the division zone. Competent-induced and non-induced *S. dysgalactiae* cells were also exposed to the fratricin in a lytic assay, revealing that the induced cells had acquired immunity.

Sammendrag

Slekten *Streptococcus* inneholder et stort utvalg av arter spredt over syv undergrupper. Den mindre kjente *Streptococcus dysgalactiae* fra pyogenic-gruppen er både et menneske- og et dyrepatogen, og er dermed delt inn i to underarter: *S. dysgalactiae* subsp. *equisimilis* (SDSE) og *S. dysgalactiae* subsp. *dysgalactiae* (SDSD), hhv. SDSD har vært i hovedfokus i denne oppgaven, hvor den har stor betydning i å forårsake bovin mastitt i kyr og septisk artritt hos lam. Dette har skapt en rekke problemer i meieriindustrien og husdyrproduksjonen, hvor antibiotika må brukes for behandling. Det har også nylig blitt observert zoonotisk sykdom i mennesker, og rapporteringer om penicillinresistente SDSE-isolater, noe som gjør *S. dysgalactiae* et viktig patogen å studere. I tillegg, ettersom arten har et komplett sett med kompetansegener og et fratrisin for kompetanse-indusert celle lysis, kan naturlig kompetanse for genetisk transformasjon være en viktig horisontal genoverføringsmekanisme for å tilegne seg gener for tilpasning og evolusjon. Følgende arbeid har derfor vært en grundig undersøkelse av fratrisin-proteinet og dets knapt karakteriserte konserverte domene. Det har tidligere blitt vist ved fluorescensmikroskop at proteinet binder til den septale regionen av målrettet celler. I denne oppgaven ble det sett, ved fusjonering av et superfoldende grønt fluorescerende protein på det konserverte domenet, at det likevel beholdt proteinets septale bindingsevne. Dette demonstrerer at domenet er ansvarlig for å dirigere proteinet til septum. Det konserverte leucin-alanin-glysin-glysin motivet og aminosyrene glutamin og tryptofan ble også identifisert ved multippelsekvensanalyse, individuelt punktmutert ved overlapp-polymerase kjedereaksjon og visualisert under fluorescensmikroskopet for å se om septalbindingsmønsteret endret seg. Ujevn binding over hele målcellen ble observert, noe som indikerer at disse høyt konserverte aminosyrene er viktig for det konserverte domenets evne til septal gjenkjennelse. Det ble også utført flere ulike protokoller for å fjerne tekoinsyrene på isolert *S. dysgalactiae* sacculi, hvor bruk av trikloreddiksyre i to uker på sacculi ble valgt til slutt. Sacculi uten tekoinsyrer viste ingen septal gjenkjennelse av fratrisinet, og det ble konkludert at disse polymerne er nødvendig for å lede proteinet mot delingssonen. Kompetent-induserte og ikke-induserte *S. dysgalactiae* celler ble også eksponert for fratrisinet i en lytisk analyse, som avslørte at de induserte cellene hadde oppnådd immunitet.

Table of contents

List of abbreviations

The important abbreviations, which were used throughout the thesis, are presented here. The page number indicates where the term and its abbreviation first appear in the text.

Abbreviation	Meaning	Page
6xHis-tag	Polyhistidine tag	51
aPBPs	Class A penicillin-binding proteins	18
APS	Ammonium persulfate	27
BHI	Brain Heart Infusion	27
CBD	Choline-binding domain	13
CbpD	Choline-binding protein D: SD = *S. dysgalactiae*, SP = *S. pneumoniae*, ST = *S. thermophilus*	12
CHAP	Cysteine, histidine-dependent amidohydrolases/peptidases	13
CIP	Alkaline Phosphatase, Calf Intestinal	28
CSP	Competence stimulating peptide	9
Cys/C	Cysteine	13
ddNTP	Dideoxynucleoside triphosphate	50
dNTP	Dideoxynucleotide triphosphate	35
dsDNA	Double-stranded DNA	11
DTT	DL-Dithiothreitol	27
EDTA	Ethylenediaminetetraacetic acid	27
Gal*N*ac	*N*-acetylgalactosamine	20
GAS	Group A *Streptococcus* (*S. pyogenes*)	1
GBS	Group B *Streptococcus* (*S. agalactiae*)	1
Glc*N*Ac	*N*-acetylglucosamine	16
Gln/Q	Glutamine	68
HGT	Horizontal gene transfer	3
IMAC	Immobilized-metal affinity chromatography	29
IAM	Iodoacetamide	13
IPTG	Isopropyl β-d-1-thiogalactopyranoside	27
LA	Lysogeny broth agar	29
LAGG	Leucine(Leu)-alanine(Ala)-glycine(Gly)-glycine	68
LB	Lysogeny broth	29
LTA	Lipoteichoic acids	20
MCS	Multiple cloning site	51
MSA	Multiple sequence alignment	68
Mur*N*Ac	*N*-acetylmuramic acid	16
OD	Optical density	60
OE-PCR	Overlap-extension PCR	40
PBPs	Penicillin-binding proteins	17
PBS	Phosphate-buffered saline	32
PCR	Polymerase chain reaction	35
Pro/P	Proline	68
SDSD	*S. dysgalactiae* subspecies *dysgalactiae*	4
SDSE	*S. dysgalactiae* subspecies *equisimilis*	4
SDS-PAGE	Sodium dodecyl sulfate–polyacrylamide gel electrophoresis	43

Continued on next page

Abbreviation	Meaning	Page
sfGFP	Superfolder green fluorescent protein	25
SH3b	Src homology 3b	13
SOC	Super optimal broth with catabolite repression	30
ssDNA	Single-stranded DNA	11
TA	Teichoic acid	13
TAE	Tris-acetate EDTA	30
TCA	Trichloroacetic acid	27
TEMED	N,N,N',N'-tetramethylethane-1,2-diamine	27
T_m	Melting temperature	36
Trp/W	Tryptophan	11
WTA	Wall teichoic acid	20
XIP	SigX-inducing peptide	10

1. Introduction

1.1 Streptococci

Although the *Streptococcus* was first recorded in 1684 by van Leeuwenhoek's drawings of microscopic images of material found between his teeth, the genus found its place in history in 1874 by Billroth who introduced its descriptive term (strepto = twisted, coccus = spherical) as well as in 1879 by Louis Pasteur who was studying puerperal fever (1, 2). The *Streptococcus* genus as seen today consists of Gram-positive ovococci bacteria, appearing in either pairs or in chains of varying lengths due to incomplete separation after cell division (3). They are non-motile, non-sporing, facultatively anaerobic although some are obligate anaerobic, catalase-negative and have complex nutritional requirements (2–4). Glucose remains the preferred energy source, though equivalent carbohydrates are also employed, where through the Embden-Meyerhof-Parnas glycolytic pathway two molecules of pyruvate are produced with simultaneous production of ATP and NADH. Streptococci are also homofermentative under oxygen limiting conditions with excess carbohydrates, allowing the bacteria to perform fermentation by reducing the pyruvate into lactic acid and regeneration of NADH. During carbohydrate limitations or increased oxygen conditions, they produce alternative fermentation products by different metabolic pathways (formate, ethanol or acetoin) (5, 6). Substrate-level phosphorylation and carbohydrate fermentation therefore primarily provide energy for the streptococci, especially since they lack oxidative phosphorylation, electron chain transport systems and retains an incomplete Krebs cycle. In contrast, however, the glucose phosphotransferase system is maintained in the genus and by using phosphoenolpyruvate derived from glycolysis as an energy source, phosphorylated carbohydrates are transported across the cellular membrane (5, 7).

Streptococcal classification has historically been based on hemolytic reactions displayed on blood agar and serological testing developed by Rebecca Lancefield in 1933 (5, 8). Hemolysis patterns are seen as either β-hemolysis (complete lysis of red blood cells), α-hemolysis (oxidation of haemoglobin to methaemoglobin causing a green patch) or γ-hemolysing meaning nonhemolytic colonies (3, 5). The Lancefield groups were applied on the β-hemolytic species only and were based on carbohydrate compositions of cell wall antigens. *Streptococcus pyogenes*, for instance, became group A *Streptococcus* (GAS) and *Streptococcus agalactiae* became group B *Streptococcus* (GBS). Sherman in 1937 made a general classification scheme based on the

Lancefield groups and other physiological characteristics to separate the genus into four divisions, “pyogenic” (β-hemolytic and Lancefield grouped), “viridans” (α-hemolytic and from oral cavity), “lactic” (later became *Lactococcus* genus) and “enterococci” (later became *Enterococcus* genus) (2, 9–11). Confusion of the streptococcal taxonomy was soon a realization as the bacteria could display various hemolytic patterns under different conditions. In addition, group-specific antigens for other species were absent or shared between distinct taxa (1, 3). 16S rRNA gene sequencing has since become the more reliable approach for identification of bacteria and has been utilized for revision of the classification system (5, 9). The pyogenic division remained as a subgroup, while the viridans division were split into five subgroups: anginosus, mitis, salivarius, bovis and mutans (Figure 1.1) (10). *Streptococcus suis* has phenotypically and genetically diverse strains where little is known, and is placed in its own group as illustrated in Figure 1.1 (12).

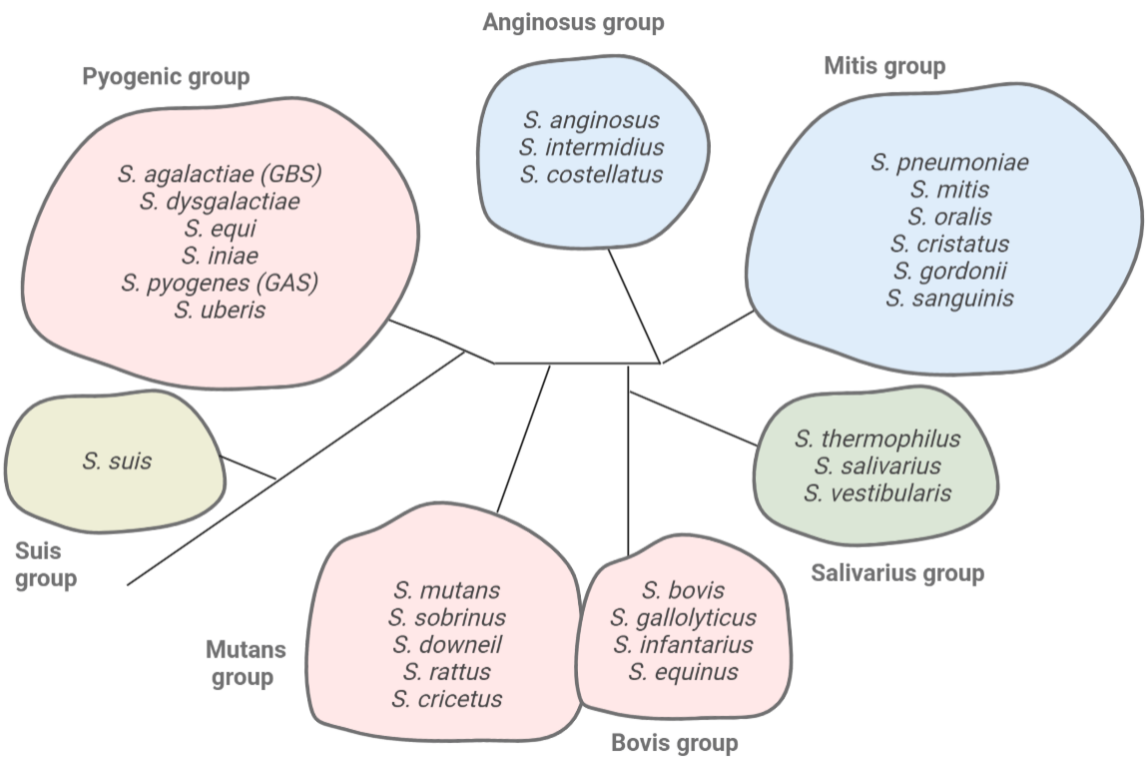

Figure 1.1: Phylogenetic relationship between the seven main streptococci subgroups as based on 16s rRNA gene sequencing. The main subgroups consists of a pyogenic group, an anginosus group, a mitis group, a salivarius group, a bovis group, a mutans group and the suis group. The pyogenic group are still Lancefield serologically-grouped based on cell wall antigen, and *S. agalactiae* is therefore given the name Group B *Streptococcus* (GBS) and *S. pyogenes* given the name Group A *Streptococcus* (GAS) (based on polysaccharides). Furthermore, anginosus group and the mitis group possess the ComCDE pathway for natural transformation, while the pyogenic group (type-II), suis group (type-III), mutans group (type-II), bovis group (type-II) and salivarius (type-I) groups possess a ComRS pathway. Figure adapted from (13) created with BioRender (14).

Streptococci are normally found as part of the microbial flora of animals and humans, with specific locations at the upper respiratory tract, alimentary and genitourinary tracts and on the skin (2, 3). Some species are restricted to a single animal or human host, while others can infect multiple hosts (10). Foods (e.g., salads, eggs and meats) and dairy products are also to be seen responsible for streptococci outbreaks (4). Furthermore, pathogenicity varies across the genus where some species are non-pathogenic commensals and others are pathogenic causing diseases ranging from subacute, acute and chronical (15). For instance, *S. pyogenes* and *Streptococcus pneumoniae* are generally associated with human disease, and have the potential to cause scarlet fever, rheumatic fever, glomerulonephritis or pneumonia to life-threatening septicemia in the case of *S. pneumoniae* (2, 3, 16). Other streptococcal species can also cause the disease bovine mastitis, which strongly plagues the dairy industry (17). A selection of antibiotics for treatment of streptococcal infections is possible, but an increasing number of antibiotic-resistant bacteria in recent decades are being observed. Antibiotic resistance can arise by spontaneous mutations or by the various mechanisms of horizontal gene transfer (HGT) for acquisition of resistance genes – where the latter is considered the most significant (18) (section 1.2). Streptococcal vaccines also present some problems, as a large number of serotypes (extracellular polysaccharide capsule or surface proteins) must be included due to the extensive genetic diversity of antigen targets (3, 19). However, there is a pneumococcal vaccine containing 23 different polysaccharide serotypes and in use in the USA today for those in high risk of serious disease (3). Yet, vaccination programs and treatment regimens face significant challenges regarding pneumococci. This species can become natural competent for genetic transformation, facilitating active uptake and genome integration of extracellular DNA. Natural competence has been shown to be the major route for spreading antibiotic resistance- and serotype genes (vaccine escape) among pneumococci (20). Most other streptococcal species also contain the genes required for natural competence, including important pathogens of the pyogenic group (21, 22). However, little is known about the contribution natural competence for genetic transformation has on gene spreading among the pyogenic streptococci.

1.1.1 *Streptococcus dysgalactiae*: a pathogen on the rise

Streptococcus dysgalactiae became recognized as a distinct species by DNA hybridization studies performed by Farrow & Collins (1984), after having lost its name in the nomenclature of Approved List of Bacterial Names in 1980 (23). They demonstrated that there was high DNA relatedness between *Streptococcus equisimilis* of Lancefield group C, the large colony-forming Lancefield

group G and β-hemolytic streptococci group L, and therefore had to belong to a single species (24, 25). In addition, Vandamme et al. (1996) showed by whole-cell protein electrophoresis patterns of several *S. dysgalactiae* strains that two distinct subpopulations of strains existed. The strains of human origin were named *S. dysgalactiae* subspecies *equisimilis* (SDSE) while the subpopulation of animal origin was named *S. dysgalactiae* subspecies *dysgalactiae* (SDSD) (23, 24). These subspecies are distinguished from each other by hemolysis and the Lancefield carbohydrate antigens, as SDSD can be α-, β- or nonhemolytic and belong to groups C and L, while SDSE are only β-hemolytic and belong to groups C and G (23).

Both SDSD and SDSE belong to the pyogenic group (Figure 1.1), and together with *S. agalactiae* and *S. pyogenes*, represent the major pathogens of this group (26). Their habitats are relatively similar, with subsp. *dysgalactiae* found in respiratory and genital tracts of various animals and subsp. *equisimilis* in the upper respiratory, gastrointestinal and female genital tracts of humans (23, 27). Furthermore, SDSD is an animal pathogen and frequently associated with bovine mastitis in cows, which is an disease that causes an inflammatory reaction of the udder tissue in their mammary glands thereby affecting milk production (28, 29). This reduces animal welfare and causes a great economic loss in the dairy industries due to lower yield and quality of the milk, as well as the need to use antibiotics (29). SDSD infections are also associated with septic arthritic outbreaks in lambs of less than four weeks old, also causing major challenges in livestock production (30). The epidemiology of SDSD is not fully known, although it is believed to be contagious. Teat lesions caused by flying insects and similar pulsed-field gel electrophoresis patterns collected from different farms also suggest that this subspecies can be found as an environmental source (28, 31). SDSE, on the other hand, is a human pathogen that can cause superficial skin infections to life-threatening toxic-like syndromes, with infections transmitted from human to human. *S. pyogenes* has a similar disease range, and it is therefore believed that this species and SDSE have near identical virulence determinants and virulence regulators. For instance, the M protein of *S. pyogenes*, encoded by the *emm* gene, allows for evasion of the host immune system by resistance to phagocytosis and adherence to human epithelial cells. The M protein is also consistently found in SDSE strains (32). SDSE is considered less virulent than *S. pyogenes*, where its infection is frequent among elderly persons and especially those with underlying medical conditions (33).

An increase in infections by SDSE has been observed worldwide, where Oppegaard et al. reported in 2014 that SDSE is the leading cause of invasive β-hemolytic disease in Western Norway (34). Within the same year, McNeilly et al. conclusively suggested that HGT and recombination has an important role of generating genetic diverse isolates of SDSE, which include giving rise to clones with increased virulence or increased fitness (35). This could thereby be an apparent link between the higher infection rate worldwide, and that SDSE isolates resistant to fluoroquinolones and penicillin have been reported (36, 37). SDSD only gives disease in animals, but there has been reports of zoonotic infective capabilities into humans. Although human diseases caused by this subspecies are rare, three cases of SDSD infections has been reported, with a fourth case recently reported in 2018 by Chennapragada et al. (38). These traits therefore make *S. dysgalactiae* (both SDSE and SDSD) a relevant and risk-imposed specie for both human and animal health. Although natural competence is important for the adaptability of pneumococci, little is known whether *S. dysgalactiae* employs natural competence to acquire genes that helps it become antibiotic-resistant and adapt to different hosts. However, considering that many streptococci, including *S. dysgalactiae*, have a complete set of competence genes, it suggests that HGT by using natural competence for genetic transformation is important for the evolution and adaptability of most streptococci (22).

1.2 Horizontal gene transfer

Horizontal gene transfer (or lateral gene transfer) involves sharing genetic material independently of reproduction events and occurs between cells with an exclusion of transmission between parent-to-offspring (39). Avery described this form of exchange in 1944 in pneumococci (40), where its significance for bacterial diversity has only become more and more apparent. HGT involves DNA transferred from donor to recipient microbe by different mechanisms, and then integration into the recipient's genome by either homologous recombination or an additive integration process (39, 41). Subsequently, HGT is an important driver of prokaryotic evolution as the horizontally acquired genetic material must confer a selective advantage in order to remain in the recipient lineage, such as adapting to a new ecological niche or bestow antibiotic resistance (39). Horizontally acquired DNA conferring an advantageous selection can then be spread quickly within a bacterial population, although some cause deleterious effects and become lost in the population over time or are even effectively neutral and depend on chance events (42).

Conjugation, transduction and transformation are the main three mechanisms of HGT (Figure 1.2). Conjugative transfer is when genetic material is transferred by cell-to-cell contact/bridge-like connection by a conjugation pilus between a donor and a recipient cell (39). Small genetic elements, such as plasmids, are often seen to be the genetic material as they can be transferred quickly, in contrast to whole chromosomes, which can take more than an hour to be transferred often resulting in the interbacterial junctions breaking down before completion (42). Transduction involves bacteriophages integrating host genetic material into their own phage genome and passing it from one bacterium to another. The genetic material can either be random portions of the host's genome (generalized transduction, lytic stage) or a specific portion of the host's DNA (specialized transduction, lysogenic stage) (39). Finally, natural transformation is the least widespread of the three HGT mechanisms among bacteria, and involves active uptake and integration of extracellular free DNA by competent bacterial cells (42). Since a significant portion of the streptococci genome consists of exogenous and mobile DNA (transposons, prophage and plasmids), HGT is thus considered the primary mechanism for creating diversity within the *Streptococcus* genus (26).

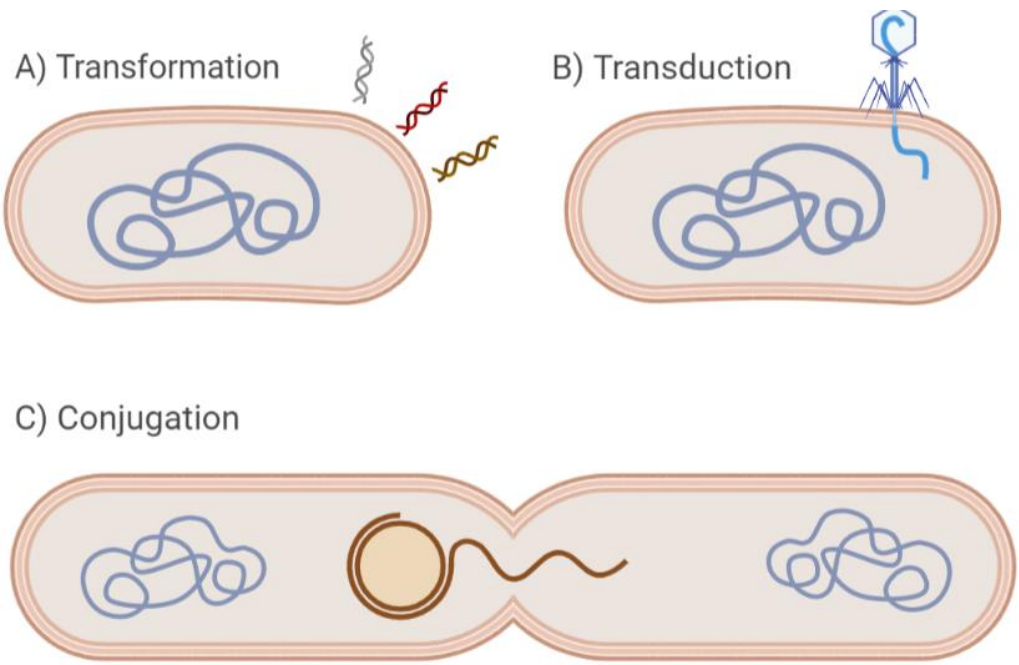

Figure 1.2: The different mechanisms of horizontal gene transfer (HGT). A) Transformation involves uptake and integration of extracellular DNA by competent bacterial cells, B) transduction involves use of viruses (bacteriophages) who integrates either a random part of the host's genome (generalized transduction) or a specific part (specialized transduction) into own phage genome, which can then be passed from one bacterium to the other. C) Conjugation involves physical contact by cell-to-cell junctions and pores through which the material can pass. Created with BioRender (14).

HGT varies greatly between streptococci, with high genetic flow found between the β-hemolytic pyogenic group of the *Streptococcus* genus. The highly similar virulence genes of *S. pyogenes* and SDSE are most likely the result of genetic exchange due to their evolutionary relationship and shared ecological niches (human hosts). Moreover, as reported by Choi et al. (2012), HGT occurs in a *S. pyogenes* to SDSE direction, with more homologous recombination than additive transfers. There is also seen more SDSE to SDSD genetic flow than the other way around, and much less SDSD to *S. pyogenes* most likely due to being in different ecological habitats (41).

1.2.1 The discovery of natural genetic transformation

Natural genetic transformation was originally discovered in *S. pneumoniae* by Frederick Griffith in 1928 (43). The "Griffith experiment" used both virulent and nonvirulent strains of *S. pneumoniae*, where it was seen that virulent colonies had a smooth appearance due to the presence of a polysaccharide capsule (S-type) whilst the nonvirulent colonies were rougher due to capsular absence (R-type). Furthermore, these R-type bacteria, as they have no protective coat, were killed by the host's immune system after injection into mice. The same was also observed for heat-killed S-type bacteria. However, by injecting both the non-lethal heat-killed S-type bacteria and live R-type into mice, it was seen that it caused a lethal infection. It was therefore believed that the polysaccharide capsule responsible for *S. pneumoniae* virulence had been passed from the heat-killed S-type into the live R-type by a chemical compound. Griffith called this observation the "transforming principle". This compound was later confirmed to be DNA, as observed by Avery, MacLeod and McCarty in 1944 (44). In 1931, Dawson & Sia also reported of a method for inducing transformation of pneumococcal types *in vitro* and thereby made it possible to better understand the mechanisms of the transformation process (45). These investigations, and further investigations through the decades, have slowly but surely decoded the transformational process and how it is regulated.

1.2.2 Induction of natural competence

Transformation is a parasexual process, which unlike conjugation and transduction, depends on the recipient cell with the necessary proteins encoded within its genome. These proteins are not expressed continuously, but rather under specific conditions to develop a specialized physiological state, a so-called competent state (46). Induction of competence in streptococci starts with expression of a competence-inducing peptide (peptide pheromone) monitored by a regulatory

system, which will be described in greater detail in section 1.2.3 (47). Although induction of competence is primarily dependent of the inducing peptide, several external signals, such as antibiotics, high phosphate concentration, bovine serum albumin and alkaline pH, indirectly cause increased concentrations of the peptide pheromone and therefore entrance to the competent state, as reported in pneumococci (46–49). For instance, in response to higher decoding rates during translation caused by the antibiotics streptomycin and kanamycin, the HtrA protease represses competence induction under normal conditions of low translational error. It does this by direct degradation of the peptide pheromone. HtrA can also degrade misfolded proteins. Therefore, when the abundance of mistranslated proteins increases, the protease no longer degrades the peptide pheromone as it is saturated by these wrongly-translated proteins (50, 51). Additionally, DNA damage also cause induction of competence, though the cause behind this is still debated. The pneumococci lacks an SOS response when their DNA is damaged by e.g., mitomycin C, causing the competence regulatory cascade to be initiated instead to coordinate a response (52).

The regulatory system is an example of quorum sensing, but in a broader sense to include population density and integration of past and current environmental factors (53). The growth phase sets the amount of cells producing the inducing peptide necessary for regulation of competence (exponential growth in pneumococci), while environmental factors modify the rate at which individual cells produce and sense the extracellular peptide pheromone, thereby up-regulating competence (46, 53).

1.2.3 Regulation of natural competence for genetic transformation

Regulation of natural competence in streptococci is performed by a two-component regulatory system, either known as the ComCDE system or the ComRS system, depending on the streptococcal species. Species having the ComCDE system do not have homologues to the ComRS proteins and vice versa and the genetic determinants are therefore different. Nevertheless, these two systems display a similar competence regulatory cascade. Both use a peptide pheromone (ComC or ComS) who's extracellular concentration is monitored by the cells. At a critical pheromone concentration, transcriptional activation of a competence regulon is triggered leading to production of all components necessary for DNA uptake and recombination (51).

The ComCDE system can be found in the anginosus and mitis groups as presented by their blue color in Figure 1.1, where *S. pneumoniae* is by far the best studied species. Details provided here

are therefore how the ComCDE system regulates competence in this model organism. A vital component of the ComCDE regulatory cascade is the CSP (competence stimulating peptide), discovered in the mid-90s by Håvarstein et al. (1995), where it coordinates competence development (47). The *comCDE* operon encodes the CSP-precursor peptide ComC, a membrane protein histidine kinase (ComD) and a response regulator (ComE). ComC has a signal sequence that is recognized by an ABC transporter composed of ComA and ComB, which causes processing of the ComC into the mature form CSP as well as secretion to the outside of the cell (47, 49, 51, 54). Meanwhile, the two-component regulatory system for signal transduction ComDE monitors the extracellular levels of CSP. At a critical level, the extracellular mature CSP binds to ComD resulting in autophosphorylation and activation of its intracellular kinase domain. It then phosphorylates the response regulator ComE at a conserved aspartate residue (49, 55). Phosphorylated ComE (ComE-P) binds to two 9-bp imperfect direct repeat motif in the promoter region of a set of genes, thereby allowing transcription of about 20 early competence genes, including the operons *comCDE* and *comAB* thus producing an auto-catalytic circuit. Among the other early *com* genes is *comX*, which encodes an alternative sigma factor (ComX) and *comM* coding for the fratricide immunity protein ComM (see section 1.3.2) (51, 56). The alternative sigma factor ComX activates the transcription of more than 80 late competence genes, necessary for fratricide (section 1.3), DNA uptake and the recombination machinery by binding to a 7-20 bp conserved region (combox) present in the promoters of these late genes (51, 57). The late-competence protein DprA is also necessary for termination of the competence period, which it does by binding to ComE-P to prevent further transcription of early genes (51).

Sequencing a large number of strains from numerous species (e.g., *S. pneumoniae*, *Streptococcus mitis*, *Streptococcus oralis*) has revealed that there exists large variations in the CSP sequences (different pherotypes), although they mostly all contain a negatively-charged N-terminal amino acid residue, positively charged C-terminal end and an arginine at the N-terminal end important for its biological function (58). Several variants of mature CSPs have been identified, although in pneumococci the vast majority of strains produce either a CSP1 or a CSP2 variant (59). Each CSP pherotype is therefore recognized by a specific and cognate ComD receptor (e.g., ComD1 or ComD2), leaving other variants unable to respond to the same receptor (58). It has also been shown that CSP1 and CSP2 contains amino acids constituting a hydrophobic patch which is essential for CSP receptor specificity (60). Pherotype switching often occurs among naturally competent

streptococci either by accumulation of point mutations or by HGT. In addition, the presence of different pherotypes of CSP allows a subset of the bacterial population to become competent and permits gene exchange within and between bacterial populations (58).

ComRS is the other regulatory system controlling competence in the pyogenic (including *S. dysgalactiae*), suis, mutans, bovis and salivarius groups. ComR is an Rgg-like transcriptional activator, while ComS is a 17 amino acid pre-peptide. ComS is secreted via an unknown transporter to form a 7-8 amino acid long mature peptide pheromone called XIP (SigX inducing peptide - SigX is a ComX homolog) (61–63). XIP is then imported back into the cell by the Ami oligopeptide transporter, where it interacts with the cytoplasmic transcriptional regulator ComR (61). The ComR is activated and binds to the early competence box (ECom box) consisting of two 9-bp inverted repeats (AACANGACAN$_4$TGTCN/ATGTTN) in the promoter regions of *comS* and *sigX*. This amplifies the production of ComS by a positive-feedback loop and induction of late competence genes by SigX, similarly as for the ComCDE system above (22, 61, 64). See Figure 1.3 for overview of the ComRS pathway.

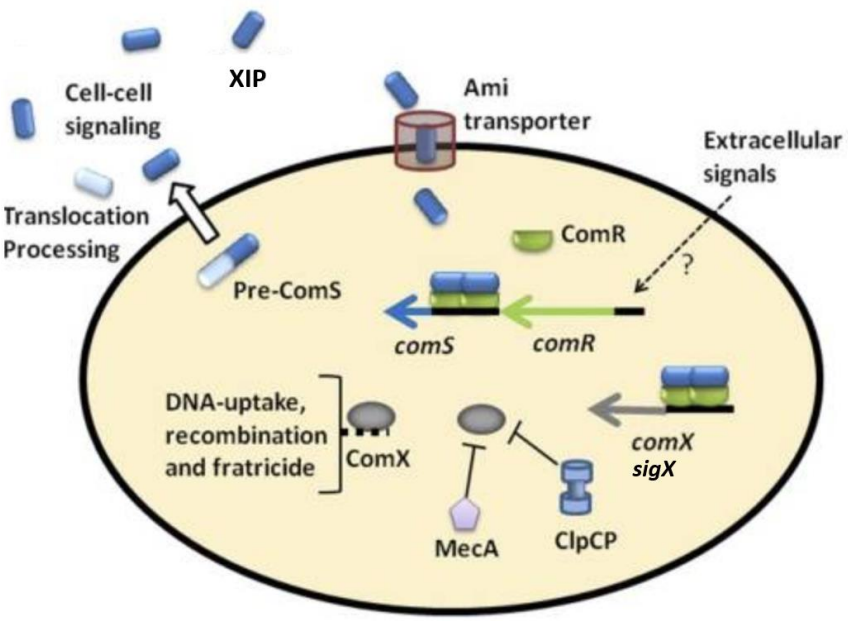

Figure 1.3: Illustrative overview of the ComRS system for regulation of natural competence as presented in *Streptococcus thermophilus*. A 17 amino acid pre-ComS peptide is transcribed from *comS* by an unknown fashion, most likely induced by extracellular signals. The precursor peptide is secreted by an unidentified transporter, processed into a mature ComS-inducing peptide (XIP) and transported back into the cell by an Ami transporter. XIP binds to the Rgg-like transcriptional regulator ComR, which activates it and cause it to bind to inverted repeats found at the promoter regions of *comS* and *sigX*. This creates a positive-feedback loop as more *comS* is transcribed, as well as expression of late competence genes. ClpC has a role in degradation of SigX under suboptimal conditions due to weak environmental stimulus. Figure acquired from (64).

The different *Streptococcal* groups can be classified into three general classes of the ComRS pathway, with the salivarius group belonging to type I, bovis, mutans and pyogenic groups of type II and the suis group of type III. This is presented in Figure 1.1, as different colors of the groups indicate different ComRS types. This classification system was implemented due to observational variations in the inverted repeats of *comS* and *sigX* promoters, and the XIP sequences (13, 22). For instance, as *S. dysgalactiae* falls in the type-II class, this species has an *comS* gene encoding a XIP with a C-terminal double-tryptophan (WW) motif. This is not observed in the type I class, although type III has a WW motif split by two residues. Subsequently, there exists different variations of the XIP signaling peptide between species, although they mostly all contain this conserved aromatic amino acid. It has also been shown that the ComR proteins can either be strict or promiscuous in recognition of the XIP messenger from other species, most likely due to a XIP-binding pocket of the ComR with two faces: a conserved face for peptide binding and a variable face for peptide pheromone specificity. Although this pocket allows binding of several XIP sequences, it is only those that satisfy correct molecular interactions with both faces and with sufficient affinity that can activate ComR. This XIP selectivity could indicate that a subset of a bacterial population becomes competent, although more work still remains to fully understand what this selectivity means for bacterial interactions in nature (13).

1.2.4 DNA uptake and homologous recombination

The extracellular free DNA is "captured" by a type IV-like pilus at the competent cell surface, and thus constitutes the primary DNA receptor (51). This transformation pilus was believed to be the case, as it was seen that a *comG* operon in Gram-positive bacteria encodes proteins homologous to proteins involved in type IV pilus assembly in Gram-negative bacteria (65). The pseudopilus delivers the double-stranded DNA (dsDNA) across the cell wall and periplasm to a membrane transport apparatus composed of ComEA, EndA, ComEC and ComFA. It is believed that depolymerization of the pseudopilus allows the DNA to travel across the cell wall (66). The dsDNA binds to the membrane receptor ComEA, while the endonuclease EndA creates a nick in one of the DNA strands. EndA also degrades the nontransforming DNA, while the transforming single-stranded DNA (ssDNA) is passed through a channel produced by ComEC in a 3’→ 5’ direction into the cytoplasm. ComFA, on the cytoplasmic side, facilitates the DNA internalization through ComEC. Homologous recombination of the ssDNA occurs by several late competence gene products, including a ssDNA-binding protein B (SsbB) which protects the DNA from degradation

by endogenous nucleases by creating a so-called eclipse complex. SsbB is then displaced by DprA, where it promotes the loading of RecA onto ssDNA. The ssDNA can thus be integrated into the chromosome by RecA-mediated homologous recombination (51, 65, 66).

1.3 The fratricide mechanism

Transformation requires uptake of homologous free DNA to facilitate chromosomal integration. Steinmoen et al. (2002) questioned if this homologous DNA is only derived from other streptococci when they die and fall apart, or if DNA is released from donor cells in a controlled and coordinated manner followed by uptake by a recipient. Following their investigations in *S. pneumoniae*, it was reported that chromosomal DNA could be released from cells during competence by a cell-lysing mechanism. It was later found that competent pneumococci express a late competence gene encoding a secreted cell wall hydrolase called CbpD (choline-binding protein D) (67). Hence, the competent pneumococci employ CbpD to actively lyse other pneumococci and close relatives, like *S. oralis* or *S. mitis*, to access homologous DNA as far-related DNA can be hazardous to the cells (68). This mechanism is called competence-induced fratricide and the lytic enzyme is called a fratricin (51). The presence of other cell wall hydrolases (LytA and LytC) in target cells of pneumococci also facilitate the lytic activity of CbpD, although these do not constitute the key components of fratricide (64). An important aspect of fratricide is also that a mixed population of competent predator cells and non-competent target cells are required. This is possible as bacteria belong to different pherogroups (section 1.2.3), meaning that one subpopulation might react with a different CSP and ComD receptor than others, causing competence to arise at different times (69). Consequently, the fratricide mechanism therefore has a large positive impact on horizontal gene transfer between streptococcal bacteria (49).

The biological function of the fratricide mechanism has suggested to be access to nutrients, to increase genetic exchange, a killing mechanism to eradicate competing bacteria or stabilize host relationships by activating innate host defenses (70). The biological function is not yet fully understood, but most data suggests that competent cells use this mechanism to obtain homologous DNA during stress. The streptococci can thereby quickly adapt to new environmental challenges as they have access to genetically different strains and species (71).

1.3.1 The choline binding protein D

The choline-binding protein D is the key protein of the fratricide mechanism in *S. pneumoniae*, where similar fratricins are also found in e.g., *S. mitis* and *S. oralis*. These fratricins have three different domains: an N-terminal CHAP (cysteine, histidine-dependent amidohydrolases/peptidases) domain, either one or two central Src homology 3b (SH3b) domains and a C-terminal choline-binding domain (CBD) with four choline-binding repeats (Figure 1.4). A signal sequence is also found at the N-terminal of the CbpD protein as shown by a computational analysis of several amino acid sequences (72). The different domains possess different biological functions, where the important CBD module binds non-covalently to the phosphorylcholine residues of teichoic acids (TAs) found at the cellular envelope of pneumococci and other close relatives (see section 1.4.2). This would therefore allow cross-species lysis as the related species contain similar recognition molecules at their cell wall (64). CBD also directs the CbpD towards the septal division zone (see section 1.4.3) (73). The catalytic CHAP domain of CbpD still has unknown enzymatic specificity, but characterized members of the CHAP family act as either endopeptidases (cleaves murein stem peptides) or as amidases (cleaves *N*-acetylmuramyl-L-Ala bonds). By this manner, and due to its homology of other cell wall hydrolases that cleaves the stem peptide, it is believed that the CbpD CHAP domain performs the same function of cleaving the peptidoglycan. An conserved active-site cysteine (C) has also been proved to be essential for its catalytic activity, as mutating it into an alanine or treatment with iodoacetamide (IAM) renders the CbpD without any lytic properties (70, 73, 74). ALE-1 is a lysostaphin homologue in *Staphylococcus aureus* which contains the SH3b domain, and where its crystal structure has revealed that it has a similar protein structure of that with its eukaryotic Src homology 3 (SH3) counterpart. It does, however, contain a unique structural feature of its own and due to low sequence similarity of the eukaryotic SH3 domain, it was believed to possess a different function (75). It was therefore proposed that the SH3b domain of CbpD domain binds to the peptidoglycan, where later findings have supported this (69, 73). It has also been suggested that the SH3b domain of CbpD specifically binds to the moieties of the glycan chains, as well as participating in a substrate-binding pocket together with the CHAP domain (73).

Experimental data has later revealed a large diversity of fratricins in different streptococcal species where the N-terminal CHAP domain is highly conserved and their C-terminal halves differs. This could perhaps indicate different cell wall binding specificities, which has most likely evolved to

limit their target range. This might be because integration of DNA from unrelated bacteria can be more harmful than beneficial to the recipient cell or cause harm to the recipient by functional incompatibility (51, 64). Figure 1.4 illustrates the domain organization of different fratricins found in streptococci. This includes, for instance, the LytF which is a functional analogue to CbpD with a C-terminal CHAP domain and is found in several streptococcal species. It has also been proposed that Zoocin A functions as a fratricin in *S. agalactiae* due to the presence of a SigX binding site upstream of the gene (64, 68).

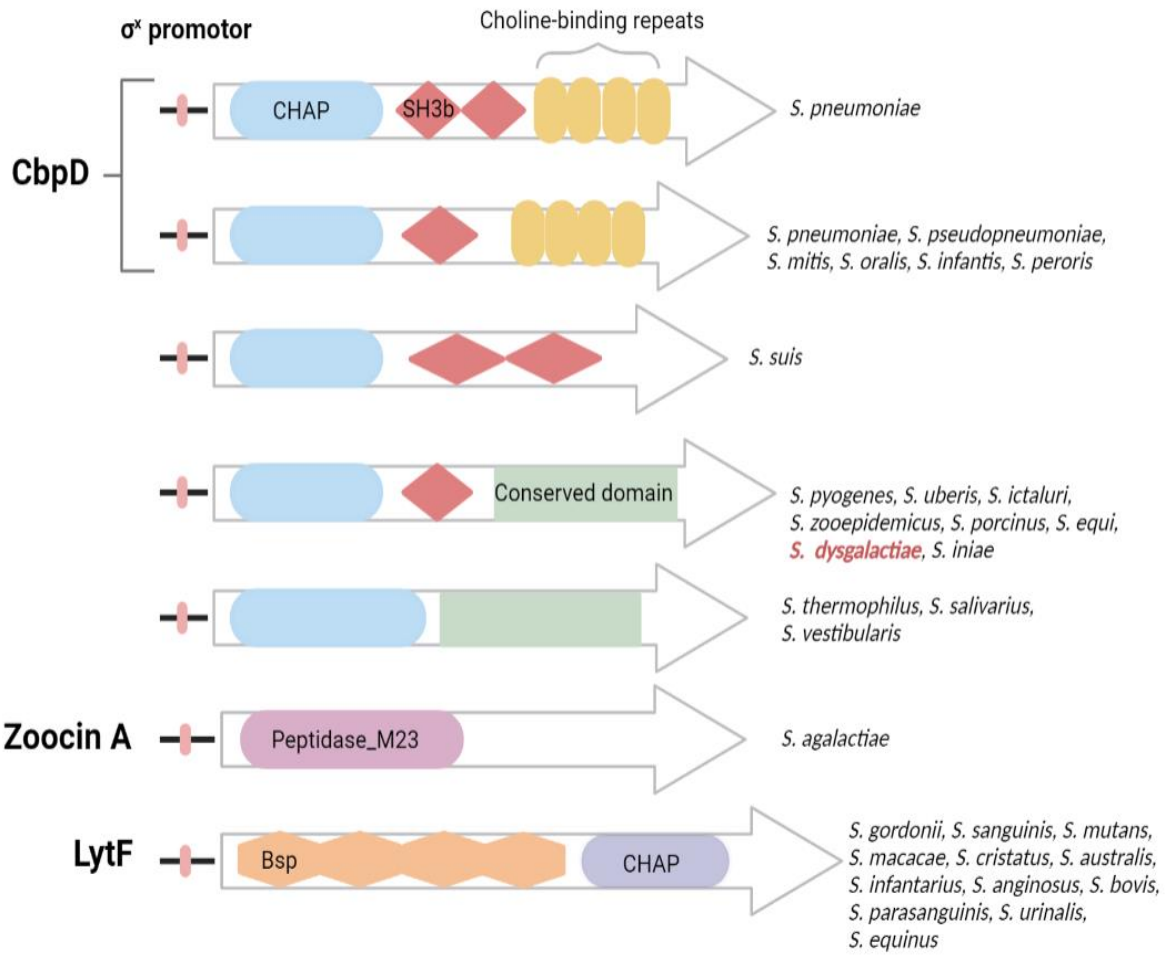

Figure 1.4: Illustrative overview of fratricins found in various streptococcal species. Common for all is the presence of a ComX/SigX promoter region upstream of the gene, as well as a catalytic CHAP domain responsible for cleaving the cell wall (except Zoocin A). The Scr homology 3b (SH3b) domain binds to the peptidoglycan and is most likely responsible for correct positioning of the CHAP domain against substrate, while the choline-binding domain (CBD) is responsible for binding to teichoic acids with choline residues at the septum of dividing cells. Some CbpDs have two SH3b domains, while others only have one or none. The CBD domain only applies for fratricins in some species, where a conserved domain can be found for other species, including *S. dysgalactiae*. Zoocin A and LytF are alternative fratricins, found in a handful of species. Figure adapted from (64) with BioRender (14).

S. dysgalactiae has also been found to have a fratricin, which confusingly is also called CbpD, although it does not contain a CBD domain. Instead it has a C-terminal conserved domain, which likely mediates binding to the cell wall of target bacterial cells (64). It is the CbpD protein from this species which is under investigation in this work, where the conserved domain is of particularly interest. Furthermore, the CbpD of *S. dysgalactiae* has a signal sequence (5'-MKKIHQLLVSGAILLSVNGAVSSVASTLNA-3'), an N-terminal CHAP domain, a single SH3b domain and the C-terminal conserved domain. The protein has a combined molecular weight of ~41 kilodalton (kDa) (~38 kDa without its signal sequence). To not confuse the various CbpD proteins and CbpD-like proteins, CbpD from *S. dysgalactiae* will henceforth be called $CbpD^{SD}$ and $CbpD^{SP}$ for *S. pneumoniae*.

1.3.2 Immunity against fratricide

The ComM protein in pneumococci is necessary to prevent the competent cells from committing suicide by their own fratricin (70). The *comM* gene has a ComE-binding promoter, meaning it is expressed as an early competence gene, where after expression constitutes a polytopic membrane protein with six to seven transmembrane segments. Furthermore, *comM* transcription peaks five minutes after competence induction, with basal expression levels after 15-20 min. This means that ComM is produced before induction of the late competence gene *cbpD*, giving the cell time to develop full immunity against the fratricin (71).

Although there are some speculations, the exact mechanism behind its immunity function is not known. It has been suggested by Straume et al. (2017) that ComM acts indirectly on CbpD, since the immunity protein is an integral membrane protein and CbpD cleaves the peptidoglycan from the outside (71). For instance, this could occur by changing the cell wall structure of peptidoglycan newly synthesized at the septal region, or modifications on the stem peptides where the catalytic CHAP domain cleaves. Changing these parts might be done as the immunity protein does not have time to modify the eternity of the cell wall, which could even be damaging to the competent cell (71, 73). Pneumococcal ComM has been the most researched immunity protein, but amino acid sequence alignments have also shown homologous proteins in other streptococcal species (71). So far, no immunity protein has been uncovered in *S. dysgalactiae*.

1.4 Streptococcal cell wall structure

The target and enzymatic substrate of fratricins is the cell wall, which functions to protect the cell from harsh environments, retain its shape and withstand turgor pressure exerted on the plasma membrane (76). The streptococci are Gram-positive bacteria and have therefore no outer membrane, but are instead surrounded by a thick peptidoglycan layer. The peptidoglycan consists of many layers, which as a whole are flexible and can thereby expand under pressure, but also contain relatively large pores for diffusion of large molecules. The cell wall also function as a scaffold for various components such as teichoic acids (section 1.4.2) and proteins, where they are needed for cell growth and division, to regulate passage of nutrients and ions, for transformation and serve as environmental sensors (77–79). The cellular envelope is often found to be the target of antibiotics due to its importance for bacterial survival: β-lactams and glycopeptides are for instance two antibiotic classes that inhibits peptidoglycan biosynthesis by various manners, ultimately causing cell lysis. As mentioned previously, antibiotic resistance is becoming more widespread possibly due to HGT events, and a push to understand the different components of the cellular envelope is becoming urgent in order to identify new cell wall targets (79, 80).

The peptidoglycan layer (also called sacculus/murein) in *Streptococcus* consists of long linear chains of alternating *N*-acetylglucosamine (Glc*N*Ac) and *N*-acetylmuramic acid (Mur*N*Ac) residues linked together by β-1 → 4 bonds, which together constitutes the glycan chain (78, 79). The glycan chains are cross-linked via stem peptides and their side-chains, where the stem peptides are composed of amino acids L-Alanine – γ-D-isoGlutamin – L-Lysine – D-Alanine – D-Alanine, although the last residue (D-Ala) is lost in the mature macromolecule. The stems are also found protruding from Mur*N*Ac by substitution of its D-lactoyl group (78, 81). The cross-linking occurs either directly or indirectly by an interpeptide bridge (Figure 1.5). The direct cross-linkage occurs via the carboxyl group of D-Ala in position 4 and the ε-amino group of L-Lys at position 3 of a stem peptide at an adjacent Mur*N*Ac at another glycan strand. The indirect cross-linking involves one of two dipeptides, seryl-alanine or alanyl-alanine, positioned between L-Lys at position 3 and D-Ala of position 4 (either L-Ala – L-Ala or L-Ser – L-Ala) (78, 81). It is believed that these variations of either linear or branched stem peptides are a result of differences in the composition of a cell wall precursor pool, caused by rate alterations of production of these compounds as reported in pneumococci (82). The general structural features of the mentioned glycan chain, stem peptides and the interpeptide bridges are commonly found for several bacteria, although a certain

degree of variation exists. It should also be taken into consideration that the peptidoglycan layer of *S. pneumoniae* has far been studied the most (78). The glycan strands of pneumococci are for instance modified in two ways: *N*-deacetylation of some Glc*N*Ac residues and O-acetylation of Mur*N*Ac residues, where both modifications provides lysozyme resistance for the cells (81). Figure 1.5 illustrates the structure of the peptidoglycan layer in *S. pneumoniae*.

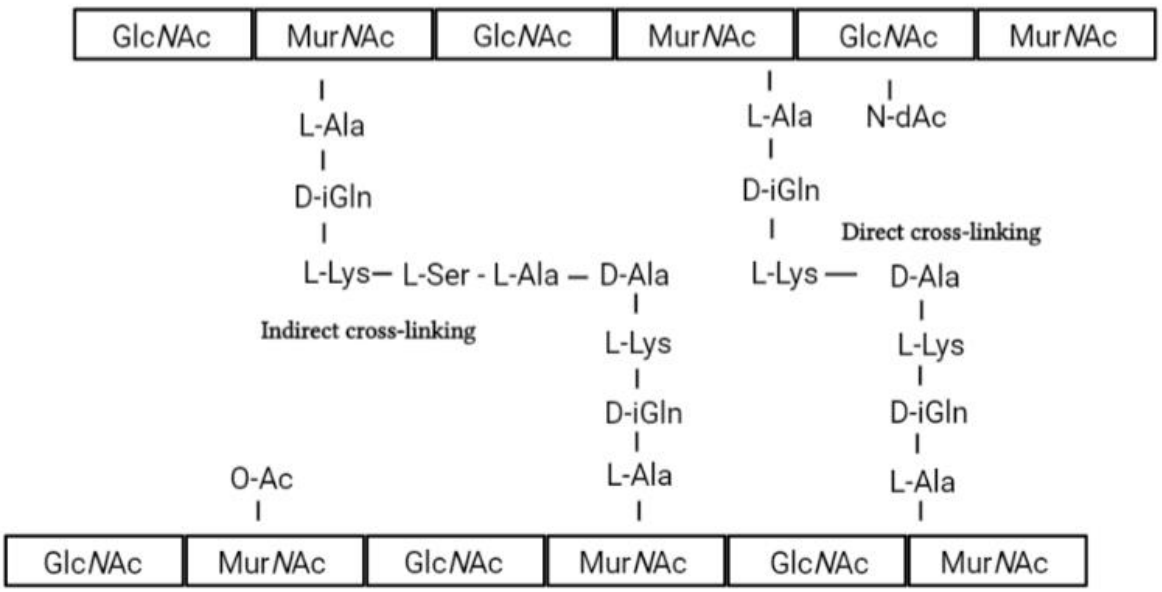

Figure 1.5: Illustrative overview of the peptidoglycan structure of pneumococci. A glycan chain consists of repeating units of *N*-acetylglucosamine (Glc*N*AC) and *N*-acetylmuramic acid (Mur*N*Ac) residues, linked together with β-1 → 4 bonds. Stem peptides (L-Ala – D-isoGlutamin – L-Lys – D-Ala – D-Ala, latter is not present in the mature peptidoglycan) attached to Mur*N*Ac, links glycan strands together to form a meshwork. It occurs by either direct cross-linkage between L-Lys of position 3 and D-Ala of position 4 of opposite strands or indirect by a interpeptide bridge either with L-Ser – L-Ala or L-Ala – L-Ala. *N*-deacetylation of some Glc*N*Ac and O-acetylation of Mur*N*Ac residues also occurs as secondary modifications. Figure adapted from (81) with BioRender (14).

1.4.1 Peptidoglycan synthesis

The biosynthesis of the peptidoglycan is a multistep process which involves several enzymes coordinated in a dynamic assembly. Different enzymatic reactions, involving synthesis of the nucleotide precursors, synthesis of lipid-linked intermediates (i.e., membrane-anchored cell wall precursor Lipid II) and polymerization reactions, occurs, with locations respectively being the cytoplasm, the inner and the outer side of the cytoplasmic membrane (83).

The membrane-bound penicillin-binding proteins (PBPs) are a large class of enzyme found on the exterior of the cytoplasmic membrane, and a part of the later stages of peptidoglycan biosynthesis (84). In *S. pneumoniae*, there are five different high molecular mass PBPs (PBP1a, PBP1b, PBP2a,

PBP2b and PBP2x). The first three are designated as class A PBPs (aPBPs) comprising bifunctional enzymes catalyzing transglycosylation and transpeptidation, while the two last are designated as class B PBPs and constitute monofunctional transpeptidases. There also exists low molecular mass PBPs, referred to as class C PBPs (PBP3), which regulates the level of peptidoglycan cross-linking (85, 86). After Lipid II is transported across the bacterial membrane (by a translocation process called flipping) to the exterior surface of the cell, the peptidoglycan can be synthesized from this precursor by polymerization of the glycan chains (transglycosylation reactions) and interconnection of the chains by peptide stem cross-linking (transpeptidation reactions) (86, 87). Importantly, the PBP2x is also necessary for septal cross wall synthesis, while PBP2b is necessary for peripheral peptidoglycan synthesis (88). PBP2x and PBP2b also have cognate partners: FtsW and the RodA proteins belong to the shape, elongation, division and sporulation (SEDS) family, and work together with PBP2x and PBP2b, respectively (Figure 1.6) (86, 89). FtsW and RodA use Lipid II as substrate to polymerize the new glycan strand, although they were initially believed to be Lipid II flippases (86, 87). It remains, however, a possibility that they possess a double role as glycosyltransferases and flippases (89).

Cell division and bacterial morphology are heavily linked to peptidoglycan metabolism, where synthesis of new cell wall (septum) at the division zone divides the mother cells into two identical daughter cells (90, 91). There are two membrane- and periplasm-spanning protein complexes necessary for peptidoglycan synthesis: the divisome responsible for cell division and septal cross wall synthesis for producing two new poles, and the elongasome for lateral insertion of peptidoglycan along the axis of the cell giving rise to its ovococci form (89, 90). The PBP2x is a key constituent of the divisome, while PBP2b is an essential member of the elongasome, as shown by the experimental results from Berg et al. (2013) (88). Cell division starts with initial inward growth of the cell wall at the cell equator which is marked by an equatorial ring (Z-ring). This equatorial ring is made by polymerization caused by the highly conserved FtsZ tubulin-like protein after its recruited to the mid-cell. On the sides of the original ring, two new equatorial rings are formed marking the future division site of the daughter cells. Moreover, the Z-ring promotes the assembly of the divisome by constituting as a scaffold. Inward septal growth continues by insertion of peptidoglycan, where the complete septum is split by peptidoglycan hydrolases to separate the daughter cells (84, 92). It is believed that the DivIVA protein is required for correct localization of the elongasome at the curve membrane region between the septal and lateral cell wall, as loss of

DivIVA localization was associated with loss of elongasome function (93). This peripheral machinery remains at the edges of the invagination of the cell membrane and ingrowth of the cell wall structure, and inserts material between the equatorial rings and the septum (84, 94). Recently, it has been reported that the septal cross-wall in Gram-positive bacteria could appear to consist of two peptidoglycan layers, one layer produced by PBP2x/FtsW and the other by aPBPs. These layers have different compositions and/or architecture as $CbpD^{SP}$ was active towards the PBP2x/FtsW-produced peptidoglycan but not towards aPBPs peptidoglycan. Based on data from Straume et al. (2020), it is believed that aPBPs work in conjunction with the divisome to synthesize a disordered peptidoglycan layer facing the cytoplasm. It also lags a few minutes behind the PBP2x/FtsW machinery (Figure 1.6) (86, 89). Furthermore, as shown in *Escherichia coli*, aPBPs might also be important for repair and maintenance of the peptidoglycan layer with similar roles suggested in Gram-positive bacteria (89).

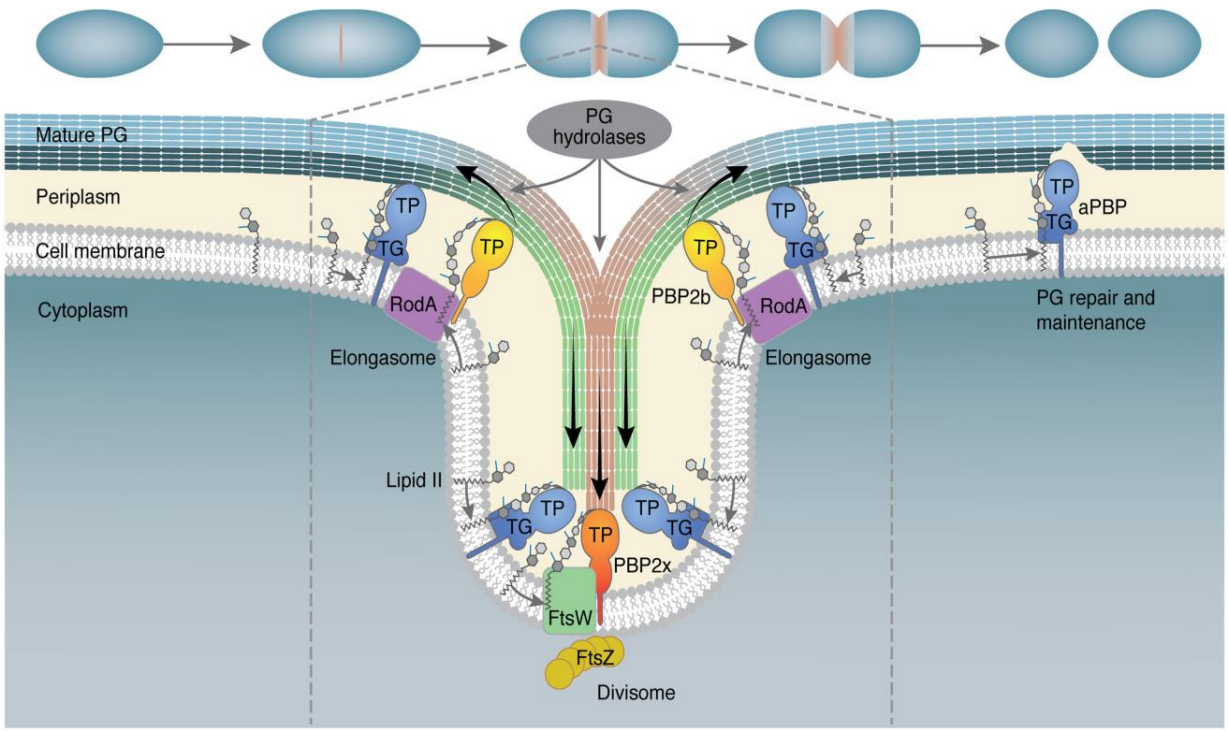

Figure 1.6: Overview of septal (divisome, PBP2x and FtsW) peptidoglycan synthesis and the peripheral (elongasome, PBP2b and RodA) peptidoglycan synthesis (brown/blue layer). The divisome is recruited mid-cell by the FtsZ (polymerizes into Z-ring), while the elongasome is believed to be recruited by a DivIVA protein. The completed septa is split by hydrolases to separate the ovococci daughter cells. A two-layered cell wall is also proposed, where class A PBPs work in conjunction with the divisome to synthesize a peptidoglycan layer facing the cytoplasm of the daughter cells (green layer). aPBPs might also have a role in repair and maintenance of the peptidoglycan layer. Figure acquired from (89).

S. pyogenes lack PBP2b and RodA necessary for the peripheral peptidoglycan synthesis, but still retains septal peptidoglycan synthesis. Therefore, this species appears more spherical and less elongated than the ovococci shape seen in pneumococci (95). *S. dysgalactiae*, on the other hand, has both the PBP2b/RodA and PBP2x/FtsW pairs, where it has been reported of amino acid substitutions in the PBP2x of penicillin-resistant SDSE isolates within and outside of its active-site motif for transpeptidase activity (37). It lacks, however, both MreC and MreD, which are two membrane proteins shown to be important for cell elongation in pneumococci. The lack of these two proteins may explain why *S. dysgalactiae* appear shorter than other ovococci (95).

1.4.2 Lipo- and wall teichoic acids

Teichoic acids are long anionic polymers which accounts for a substantial fraction of the cell wall mass (~60 %) of Gram-positive bacteria (76). The TAs polymers contribute to a variety of processes, including resistance to environmental stresses (e.g., heat, low osmolarity), trafficking of ions and nutrients, antibiotic resistance, cation homeostasis and acts as receptors for phage particles. Cell surface proteins involved in interactions with host factors can also be found noncovalently bound to the TAs (77, 96, 97). For many Gram-positive bacteria (including most of the streptococcal species), the TAs consists of repeating units of polyglycerol phosphates or polyribitol phosphates of variable length that can carry either D-alanine residues, which reduces the negative charge of the cell envelope, or glycosyl residues (77). There also exists two main classes of TAs: the lipoteichoic acids (LTAs) and the wall teichoic acids (WTAs). The LTAs are anchored to the cytoplasmic membrane by possessing a glycolipid moiety functioning as an anchor, while the WTAs are covalently attached to the Mur*N*Ac residues of the peptidoglycan layer (98) (Figure 1.7A). Insertion of newly synthesized TAs into the cell wall also occurs at the mid-cell, where the septal cross-wall is formed (99).

S. pneumoniae is unique in that the WTA and LTA are most likely produced in the same biosynthetic pathway as they have similar repeating units and length distribution (97). The pneumococci TAs are a tetrasaccharide-ribitol structure, where the repeated unit of the main chain is composed of a 2-acetamido-4-amino-2,4,6,-trideoxygalactose (AATGal) sugar, D-glucose (Glc), ribitol-5-phosphate (Rib-P) and two Gal*N*ac (*N*-acetylgalactosamine) residues each decorated with phosphorylcholine (Figure 1.7B). The degree of phosphocholine substitution varies. These choline decorated TAs, as previously mentioned, are therefore anchor sites for choline-binding proteins

(97, 100). It has also been reported of a lack of D-alanylation in *S. pneumoniae*, which contrasts the TAs of many other Gram-positive bacteria, though the *dlt* operon which encodes proteins necessary for D-alanine substitution has been found in both clinical and laboratory pneumococcal strains. Kovacs et al. (2006) reported that preparations of TAs could cause loss of D-alanyl esters and that pneumococci TAs therefore contains these substitutions (100).

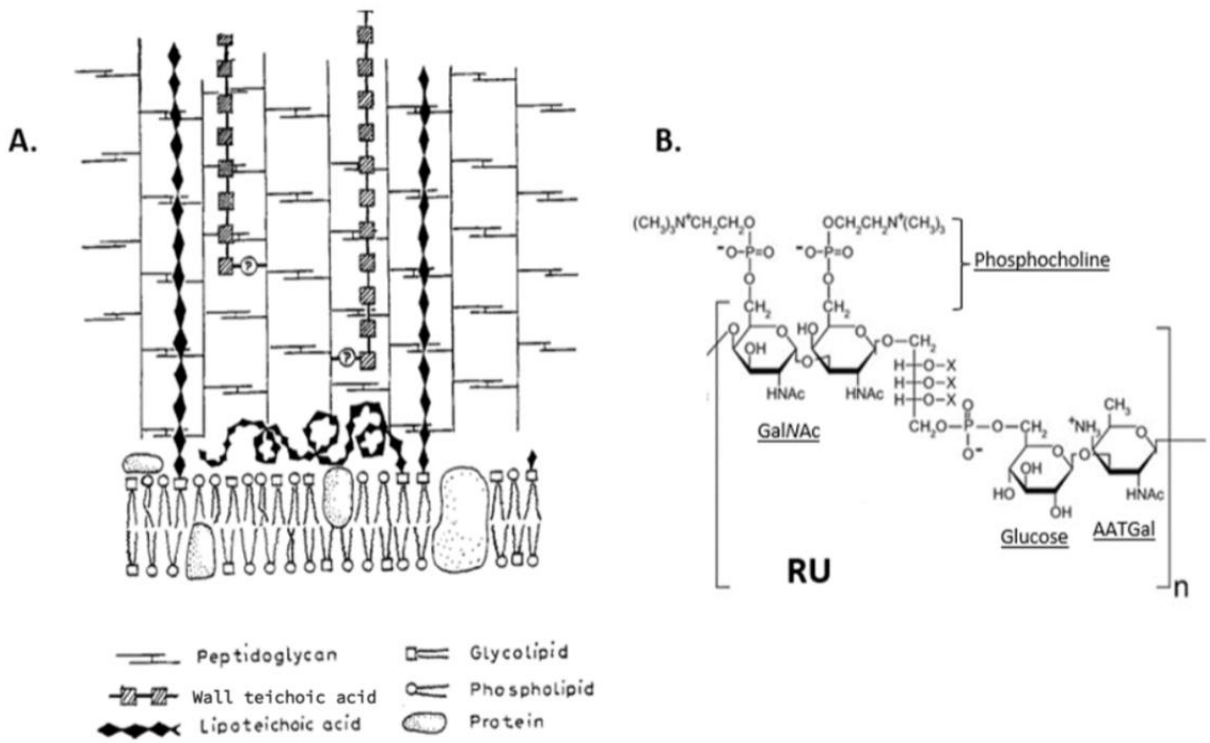

Figure 1.7: A. Overview of pneumococci lipoteichoic acids (LTAs) and the wall teichoic acids (WTAs) anchored respectively either to the cytoplasmic membrane or to Mur*N*Ac residues at the peptidoglycan layer. Some cell surface proteins can also be found anchored to the TAs. This sketch is provided by Werner Fischer (1997) (98). **B**. Illustrates pneumococcal repeating units (RU) of both LTAs and WTAs main chain, with a 2-acetamido-4-amino-2,4,6-trideoxygaalactose sugar (AATGal), glucose, ribitol phosphate (Rib-P) and two *N*-acetylgalactosamine (Gal*N*Ac) with phosphorylcholine. Figure obtained from (97). Both figures adapted with BioRender (14).

Most details obtained on streptococcal cell wall structure comes from studies on *S. pneumoniae*. Therefore, little information is available on the cell wall structure of *S. dysgalactiae*, which is the organism that has been studied in the present work. It has however been reported by Czabanska et al. (2012) that the LTA backbone of *S. dysgalactiae* has a poly(sn-glycerol-1-phosphate) (GroP) structure with 37% substitutions at O-2 with D-Ala. A kojibiose-diacylglycerol lipid anchor was also identified (101) (Figure 1.8). In addition, the species has a group C carbohydrate, where the repeating unit of the backbone is rhamnose with Gal*N*Ac side-chains (102). More research has been

performed on the related species *S. pyogenes* of which belong to the same pyogenic group as *S. dysgalactiae*. The cell walls are composed of a peptidoglycan network identical to the peptidoglycan description in section 1.4, with additional type-specific M, R and T proteins. The stem peptides are also identical of what has been discussed, although a ratio of 10:1 lysine to hydroxylysine has been found (78, 103). Furthermore, an indirect linkage sites necessary for cross-linking occurs by an interpeptide bridge of the dipeptide L-alanyl-L-alanine (103). *S. pyogenes* as GAS has a glycopolymer, instead of expression of a classical WTA, composed of a rhamnose polysaccharide backbone with Glc*N*Ac side-chain modifications on every α-1,2–linked rhamnose (102). On the other hand, the LTAs have polyglycerol phosphate repeating units, where sugars and D-alanine can be substituents (104).

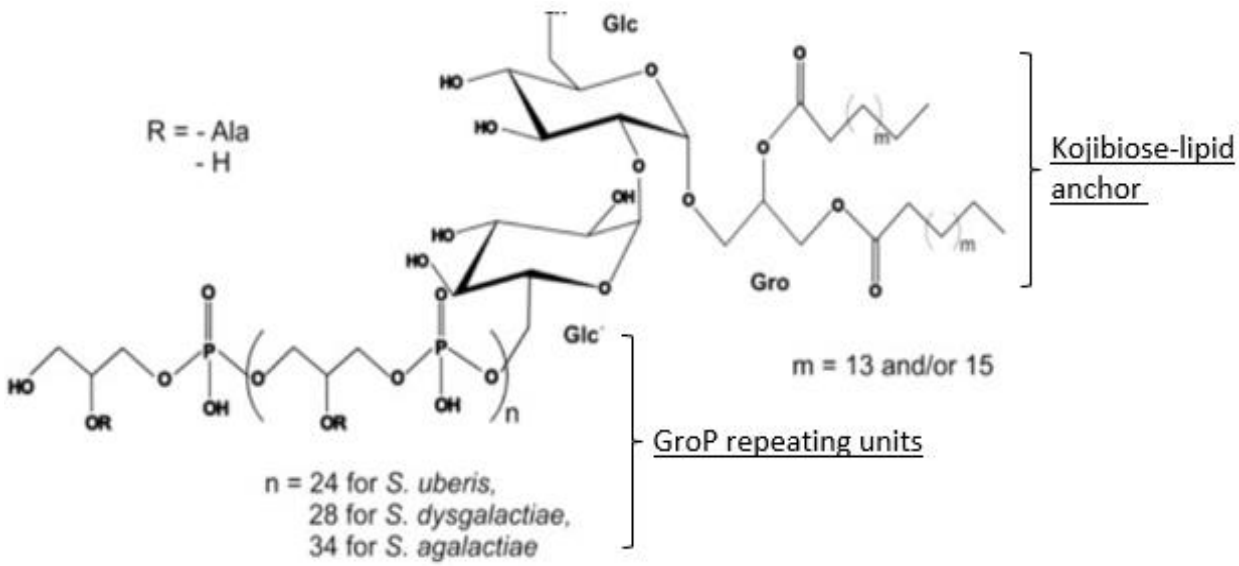

Figure 1.8: The lipoteichoic acids from *S. dysgalactiae* strain 2023. Its backbone chain consists of a poly(sn-glycerol-1-phosphate) (GroP) repeating units with 37% substitution at O-2 with D-alanine. It also has a kojibiose-containing anchor structure linked with diacylglycerol with the following structure α-D-Gl*cp*-(1→2)-α-D-Gl*cp*-(1→3)-1,2-diacyl-*sn*-glycerol. It also has 28 repeating units of GroP within their LTA. Figure acquired from (101) and adapted with BioRender (14).

1.4.3 CbpD^{SP} attacks the septum of target cells

It has been shown that pneumococci treated with CbpD^{SP} specifically cause rupture at the cell septum. CbpD^{SP} uses its CBD to bind noncovalently to the choline-moieties of the TAs located at the septum region of target cells, where it cleaves the septal cross wall (synthesized by the divisome, PBP2x/FtsW) (73). How CbpD^{SP} discriminates between TA in newly synthesized cell walls and mature cell walls is not known. Similarly, it has been shown that LytF from *S. gordonii*

also binds to the septum of target cells (68). A previous master's degree thesis performed in the Molecular Microbiology group explored the localization of $CbpD^{SD}$ from *S. dysgalactiae,* and showed that also this fratricin localized to the septum of target cells (105). In summary, this strongly suggests that the newly synthesized peptidoglycan layer serves as a substrate for these lytic enzymes and that the domains, other than the CHAP domain, are important for recognizing target cells and guiding the fratricins to the division zone. Evidently, TA is critical for septal localization of $CbpD^{SP}$, but whether TA is important for target recognition and septum localization of LytF or $CbpD^{SD}$ is not known. Considering that the differences in TA structure between streptococcal species correlates with production of a fratricin with different target recognition domains, it is likely that TA is the key recognition molecule also for these. However, experimental evidence for this is lacking.

1.5 Aim of study

S. dysgalactiae is both an animal pathogen (SDSD) with potential to cause bovine mastitis, as well as a human pathogen (SDSE) with potential to cause life-threatening toxic-like syndromes. An increase in cases of infections by this species has been observed worldwide, with penicillin-resistant bacteria and zoonotic diseases also reported. This combination therefore has the potential to cause serious issues in the future if more patients are hospitalized and cannot be properly treated with antibiotics due to the emergence of resistant strains. In light of the situation with pneumococci, where natural transformation and fratricide are crucial for the spread of resistance genes and for vaccine escape, it is critical to determine the potential of similar scenarios for this "new" pathogen *S. dysgalactiae*. Since *S. dysgalactiae* has a $CbpD^{SD}$ fratricin with lytic activity that appears to be competence regulated (having a SigX binding site in the $P_{cbpD}{}^{SD}$ promoter), fratricide could potentially be an important mechanism for this bacterium to gain new traits. The current work was therefore initiated to further characterize the *S. dysgalactiae* fratricin, where it is known that $CbpD^{SD}$ causes lysis of target cells by its CHAP domain. Little is known about its conserved domain, although it is believed to be involved in target cell recognition. Understanding the biochemical properties of $CbpD^{SD}$ allows for better understanding of *S. dysgalactiae* evolution and adaptability, as not much is known about its importance in becoming natural competent for genetic transformation. This research has therefore focused on the following:

- Overexpress and purify different $CbpD^{SD}$ versions from *S. dysgalactiae* subsp. *dysgalactiae* strain MA201
- Study whether the conserved domain of $CbpD^{SD}$ alone accounts for septum binding
- Identify conserved amino acids in the conserved domain important for its function
- Study if $CbpD^{SD}$ binds to teichoic acids
- Check if *S. dysgalactiae* strains are immune against $CbpD^{SD}$ activity

2. Materials

2.1 Vectors and bacterial strains

Table 2.1: List of vectors used in this work, with a short description, variants and source of vectors also included.

Vector	Description	Variants[a]	Source
pRSET A	High-level expression vector for recombinant proteins	pRSET-A-*cbpD*^{SD}, amp^{R} pRSET-A-*sfgfp*^{b}-*cbpD*^{SD}, amp^{R}	Lab collection
pRSET A	High-level expression vector for recombinant proteins	pRSET-A-*cbpD*^{G302A,G303A}, amp^{R} pRSET-A-*cbpD*^{C65A}, amp^{R} pRSET-A-*sfgfp-cbpD*$^{\Delta SH3b}$, amp^{R} pRSET-A-*sfgfp-cbpD*$^{\Delta SH3b,\ L300S,\ A301S}$, amp^{R} pRSET-A-*sfgfp-cbpD*$^{\Delta SH3b,\ G302A,G303A}$, amp^{R} pRSET-A-*sfgfp-cbpD*$^{\Delta SH3b,Q329A}$, amp^{R} pRSET-A-*sfgfp-cbpD*$^{\Delta SH3b,W365A}$, amp^{R}	This work

a. ampR; ampicillin resistance
b. sfGFP = superfolder green fluorescent protein

Table 2.2: List of *Escherichia coli* strains used in this work, with characteristics and source of strain also included.

Escherichia coli strain	Characteristics	Source
DH5α	Parental strain: pRSET-A-CbpD^{SD} vector, amp^{R}	Invitrogen™
BL21	Parental strain: pRSET-A-CbpD^{SD} vector, amp^{R}	Invitrogen™
JA1	DH5α, pRSET-A-CbpD^{SD}, amp^{R}	Lab collection
JA2	BL21, pRSET-A-CbpD^{SD}, amp^{R}	Lab collection
JA6	DH5α, pRSET-A-sfGFP-CbpD^{SD}, amp^{R}	Lab collection
JA7	BL21, pRSET-A-sfGFP-CbpD^{SD}, amp^{R}	Lab collection
JA8[a]	BL21, pRSET-A-sfGFP-CbpD^{SD}, amp^{R}	This work
JA9	DH5α, pRSET-A-sfGFP-CbpD$^{\Delta SH3b}$, amp^{R}	This work
JA10	BL21, pRSET-A-sfGFP-CbpD$^{\Delta SH3b}$, amp^{R}	This work
JA11	DH5α, pRSET-A-sfGFP-CbpD$^{\Delta SH3b,\ G302A,G303A}$, amp^{R}	This work
JA12	DH5α, pRSET-A-sfGFP-CbpD$^{\Delta SH3b,\ L300S,\ A301S}$, amp^{R}	This work
JA13	DH5α, pRSET-A-sfGFP-CbpD$^{\Delta SH3b,W365A}$, amp^{R}	This work
JA14	DH5α, pRSET-A-sfGFP-CbpD$^{\Delta SH3b,Q329A}$, amp^{R}	This work
JA15	BL21, pRSET-A-sfGFP-CbpD$^{\Delta SH3b,\ G302A,G303A}$, amp^{R}	This work
JA16	BL21, pRSET-A-sfGFP-CbpD$^{\Delta SH3b,\ L300S,\ A301S}$, amp^{R}	This work
JA17	BL21, pRSET-A-sfGFP-CbpD$^{\Delta SH3b,W365A}$, amp^{R}	This work
JA18	BL21, pRSET-A-sfGFP-CbpD$^{\Delta SH3b,Q329A}$, amp^{R}	This work
JA19	DH5α, pRSET-A-CbpD^{G302A,G303A}, amp^{R}	This work
JA20	BL21, pRSET-A-CbpD^{G302A,G303A}, amp^{R}	This work
JA21	DH5α, pRSET-A-CbpD^{C65A}, amp^{R}	This work
JA22	BL21, pRSET-A-CbpD^{C65A}, amp^{R}	This work
JA23[b]	BL21, pRSET-A-CbpD^{C65A}, amp^{R}	This work

a. transformed vector from JA6 into BL21 again for verification
b. transformed vector from JA21 into BL21 again for verification

Table 2.3: List of *Streptococcus dysgalactiae* subspecies *dysgalactiae* strains used in this work, with characteristics and source of strain also included.

SDSD strain	Characteristics	Source
MA201	Parental strain: truncated ComR, isolated from cow, wt	Norwegian Veterinary Institute
JA5	MA201, wt	This work
MM352	Strain 015, Δc*bpD*::*spectinomycin*, isolated from cow	Lab collections
MM356	Strain 047, isolated from cow, wt	Lab collections

2.2 Primers

Table 2.4: Overview of the primers used in this work, with primer sequence (5'→ 3') and source also included.

Primer name	Sequence 5'→ 3'	Source
Primers for validation of c*bpD*^{SD} and for use in sequencing		
pRSET (F)[a]	AATACGACTCACTATAGGGAGA	Invitrogen™
pRSET (R)	CTAGTTATTGCTCAGCGGT	Invitrogen™
Construction of *sfgfp-cbpD*$^{\Delta SH3b}$		
JAA1 (F)	AAAAAGACGGGGCAAAAAACAC	This work
JAA2 (R)	GTGTTTTTTGCCCCGTCTTTTTTACTGCTTCTTTAGAAACCTGTTG	This work
Construction of *sfgfp-cbpD*$^{\Delta SH3b,\ G302A,G303A}$ and c*bpD*^{G302A,G303A}		
JAA3 (F)	GCTGCTGGCGCTACTTCTCTTAACTG	This work
JAA4 (R)	CAGTTAAGAGAAGTAGCGCCAGCAGCAGCTAGCTCCTCGCTAACTAATAG	This work
Construction of *sfgfp-cbpD*$^{\Delta SH3b,\ L300S,\ A301S}$		
JAA5 (F)	AGTAGTGGTGGGGGCGCTACTTCTC	This work
JAA6 (R)	CAGTTAAGAGAAGTAGCGCCCCCACCACTACTCTCCTCGCTAACTAATAG	This work
Construction of *sfgfp-cbpD*$^{\Delta SH3b,W365A}$		
JAA7 (F)	GCTTTAACTGCTGATAAAGCCTCTAAATTG	This work
JAA8 (R)	TAGAGGCTTTATCAGCAGTTAAAGCTGTTCCACGTGATCCAATC	This work
Construction of *sfgfp-cbpD*$^{\Delta SH3b,Q329A}$		
JAA9 (F)	GCTATTTTACAGGCCGGTGAGTTTTTTG	This work
JAA10 (R)	AAACTCACCGGCCTGTAAAATAGCATTTCCTGCTTTTACTCCTTTAC	This work
Construction of c*bpD*^{C65A}		
JAA11 (F)	GCCACATCATTTGTCGCTTTCCGTC	This work
JAA12 (R)	GACGGAAAGCGACAAATGATGTGGCCTGACGAACATACATATTCCAG	This work

a. (F) – forward primer, (R) – reverse primer

2.3 Chemicals

Table 2.5: Overview of the various chemicals used in this work, with product number and supplier also listed.

Chemical[a]	Product number	Supplier
Acetic acid, CH3COOH	100063	Merck
Acrylamide 4x (40%), C_3H_5NO	A0962.1000	VWR
Agarose	15510-027	Invitrogen™
Agar powder	20768.292	VWR
Ammonium persulfate (APS), $(NH_4)_2S_2O_8$	1610700	Bio-Rad
Bacto™ Yeast Extract	212750	BD Diagnostics
Brain Heart Infusion (BHI) Broth	CM1135	OXOID
Bromophenol blue, $C_{19}H_{10}Br_4O_5S$	115-39-9	Sigma-Aldrich
Calcium chloride, $CaCl_2$	102378	Merck
Coomassie brilliant blue, $C_{37}H_{34}N_2Na_2O_9S_3$	20278	ThermoFisher Scientific™
Disodium phosphate (DSP), Na_2HPO_4	106586	Merck
DL-Dithiothreitol (DTT), $C_4H_{10}O_2S_2$	D0632-10G	Sigma-Aldrich
Ethanol (96%), C_2H_5OH	20824	VWR International
Ethylenediaminetetraacetic acid (EDTA), $C_{10}H_{16}N_2O_8$	108418	Merck
Glass beads, acid washed ≤ 106 µm (−140 U.S. sieve)	G4649-100G	Sigma-Aldrich
Glycerol 85%, $C_3H_8O_3$	49781-1L	Sigma-Aldrich
Hydrochloric acid fuming 37%, HCl	100317	Merck
2-iodineacetamid, C_2H_4INO	8047440025	Merck
Imidazole, $C_3H_4N_2$	I2399	Sigma-Aldrich
Isopropyl β-d-1-thiogalactopyranoside (IPTG), $C_9H_{18}O_5S$	A1008.0005	PanReac AppliChem
Lithium chloride, LiCl	L9650	Sigma-Aldrich
Magnesium sulfate, $MgSO_4$	2312982	Sigma-Aldrich
Monopotassium phosphate (MKP), KH_2PO_4	1.04873.1000	Merck
peqGREEN	PEQL37-501	Saveen Werner
Potassium chloride, KCl	1.04936.1000	Merck
Sodium chloride, NaCl	27810.295	VWR
Sodium dodecyl sulfate (SDS), $CH_3(CH_2)_{11}OSO_3Na$	05030	Fluka®
Sodium hydroxide, NaOH	1.06498.1000	Merck
Sucrose, $C_{12}H_{22}O_{11}$	102745C	Sigma-Aldrich
N,N,N',N'-tetramethylethane-1,2-diamine (TEMED), $C_6H_{16}N_2$	1610800	Bio-Rad
Trichloroacetic acid (TCA), $C_2HCl_3O_2$	91699	Fluka®
Trizma® base $C_4H_{11}NO_3$	SLBZ4871	Sigma-Aldrich
Tween® 20, $C_{58}H_{114}O_{26}$	1706531	Bio-Rad

a. Chemicals for C-medium not included

2.4 Antibiotics

Table 2.6: Overview of the antibiotic used in this work, with stock solution, product number and supplier also listed.

Antibiotic	Stock solution	Product number	Supplier
Ampicillin	100 mg/mL	A9251	Sigma-Aldrich

2.5 Enzymes, stains, size markers and nucleotides

Table 2.7: Overview of the various enzymes, stains, size markers and nucleotides used in this work, with product number and supplier also listed.

Compound	Product number	Supplier
1 kb DNA ladder	N3232	New England BioLabs®
Alkaline Phosphatase, Calf Intestinal (CIP)	MO525S	New England BioLabs®
Broad Range Protein Ladder	P7719S	New England BioLabs®
Deoxyribonuclease I from bovine pancreas	DN25	Sigma-Aldrich
dNTPs (dATP, dGTP, dCTP, dTTP)	N0447	New England Biolabs®
HindIII restriction enzyme	R0104S	New England BioLabs®
Lysozyme human	LI667	Sigma-Aldrich
NdeI restriction enzyme	R0111S	New England BioLabs®
Phusion® High-Fidelity DNA polymerase	M0530L	New England BioLabs®
Red*Taq* 2x Master mix 1.5 mM $MgCl_2$	52000300-1250	VWR Life Science
RNase A from bovine pancreas	R6513	Sigma-Aldrich
SYTOX™ Green Nucleic Acid Stain	S7020	Invitrogen™
Trypsin from porcine pancreas	93615	Sigma-Aldrich
Quick Ligase	M2200L	New England BioLabs®

2.6 Inducer peptides

Table 2.8: Overview of the inducer peptide used in this work, its amino acid sequence (N-terminal → C-terminal direction), stock solution and supplier are also included.

Inducer peptide	Amino acid sequence (N→ C)	Stock solution	Source
ComS2	QVDWWRL	0.01 mg/mL	Research Genetics, Inc

2.7 Kits

Table 2.9: Overview of the various kits used in this work, with product name and supplier also included.

Kit	Product number	Supplier
96-well polystyrene microtiter plates	82.1581.001	Sarstedt
E.Z.N.A. ® Plasmid DNA Mini Kit I	D6943-02	Omega Bio-Tek
NucleoSpin® Gel and PCR Clean-up	740609.250	Macherey-Nagel®
Protino Ni-TED 150 kit in Mini format	745100.10	Macherey-Nagel®
Slide-A-Lyzer™ Dialysis Cassette G2	87730	ThermoFisher Scientific™
Spectra/Por molecularporous membrane MWCO 12-14.000	17127	Spectrum®

2.8 Equipment

Table 2.10: Overview of the equipment used in this work, with model and supplier also listed.

Equipment	Model	Supplier
Agarose electrophoresis system	Mini-Sub Cell GT cell	Bio-Rad
Chromatopgrahpy system	ÄKTA pure 25L	Cytiva
Electrophoresis chamber for agarose gel	Mini-Sub Cell GT Cell	Bio-Rad
Gel imager I	GelDoc-1000	Bio-Rad
Gel imager II	Azure Imager c400	Azure Biosystems
Grinder and lysis system	FastPrep®-24	MP™ Biomedicals
Homogenesing system	"Pressure cell" homogenizer SPCH-EP	Stansted®
IMAC[a] column	HisTrap HP precharged with nickel	Cytiva
Microplate reader	Hidex Sense	Hidex Oy
Microscope	LSM700	Zeiss
Microscope camera	ORCA-Flash4.0 V3 Digital CMOS camera	Hamamatsu Photonics K.K.
PCR machine	ProFlex™ PCR system	Applied Biosystems™
SDS-PAGE system	Mini-PROTEAN® Tetra Handcast Systems	Bio-Rad
Spectrophotometer I	NanoDrop™ 2000	ThermoFisher Scientific™
Spectrophotometer II	GENESYS™ 30 Visible Spectrophotometer	ThermoFisher Scientific™
Spectrophotometer III	Pharmacia type Novaspec II	Pharmacia

a. IMAC = immobilized metal affinity chromatography

2.9 Growth medium, buffers and solutions

Table 2.11: Overview of premade buffers used in this work, with product number and supplier also included.

Buffer	Product number	Supplier
5x Phusion High-Fidelity reaction buffer	B0518S	New England BioLabs®
2x Quick Ligase reaction buffer	B2200S	New England BioLabs®
NEBuffer™ 2.1 for restriction cutting	B7202S	New England BioLabs®

2.9.1 Growth media

Lysogeny Broth (LB) medium

10 g/L NaCl, 10 g/L tryptone, 5 g/L yeast extract

Autoclaved.

LB agar (LA)

10 g/L NaCl, 10 g/L tryptone, 5 g/L yeast extract, 15 g/L agar powder

Autoclaved, cooled down to 55 ℃ and appropriate antibiotics added.

Super optimal broth with catabolite repression (SOC) medium

20 g/L tryptone, 5 g/L yeast extract, 10 mM sodium chloride, 2.5 mM potassium chloride
Autoclaved, with 20 mM magnesium chloride and 20 mM glucose added afterwards.

C-medium

150 µL 0.4 mM $MnCl_2$, 1.5 mL 20% (w/v) glucose, 3.75 mL ADAMS III, 110 µL 3% (w/v) glutamine, 2.25 mL 2% (w/v) Na pyruvate, 95 mL 1.5 M sucrose, 1.5 mL 2 mg/mL uridine adenosine, 1.5 mL 8% (w/v) albumin/BSA, 3.75 mL yeast extract
Added to 150 mL pre-C medium and stored fresh at 4 °C. See appendix 1 for pre-C medium, ADAMS I, ADAMS II and ADAMS III recipes.

2.9.2 Buffers and solutions for agarose gel electrophoresis

6x DNA loading buffer

10 mM Tris-HCl (pH 8.0), 1 mM EDTA, 40% (w/v) sucrose, 0.01% (w/v) bromophenol blue.

50x Tris-acetate EDTA (TAE buffer)

242 g Tris base, 57.1 mL acetic acid, 100 mL 0.05 M EDTA (pH 8.0)
Volume adjusted to 1 L with dH_2O.

1% agarose gel

0.5 g agar, 50 mL 1x TAE buffer, 1 µL peqGREEN
Heated until agar was dissolved and then added peqGREEN.

1 kb DNA ladder

10 µg 1kb ladder, 75 µL 10x loading buffer
Adjusted to 750 µL using dH_2O and stored at 4 °C.

2.9.3 Buffers and solutions for sodium dodecyl sulfate-polyacrylamide gel electrophoresis

2x SDS sample buffer

0.125 M Tris-HCl (pH 6.8), 4% (w/v) SDS, 0.2 M DTT, 20% (v/v) glycerol, 0.01% (w/v) bromophenol blue
Stored at 4 °C.

10x SDS running buffer

144 g glycine, 30.2 g Tris base, 1% (w/v) SDS

dH_2O was added to a final volume of 1 L.

12% Separation gel

4.3 mL dH_2O, 2.5 mL 1.5M Tris-HCl (pH 8.8), 0.1 mL 10% (w/v) SDS, 3.0 mL 40% acrylamide + 0.8% `bis`-acrylamide, 0.1 mL 10% (w/v) APS, 0.005 mL TEMED

Mixed all reactants together, except from APS and TEMED which was added last, right before casting the gel.

4% Stacking gel

3.15 mL dH_2O, 1.25 mL 0.5M Tris-HCl (pH 6.8), 0.05 mL 10% (w/v) SDS, 0.5 mL 40% acrylamide + 0.8% `bis`-acrylamide, 50 µL bromophenol blue, 0.05 mL 10% (w/v) APS, 0.005 mL TEMED

Mixed all reactants together, except from APS and TEMED which was added last, right before casting the gel.

Coomassie blue stain

0.1 g 0.1% (w/v) Coomassie Brilliant Blue, 250 mL 50% (v/v) ethanol, 37.5 mL 7.5% (v/v) acetic acid

dH_2O was added to a final volume of 500 mL.

Destainer solution

50 mL (v/v) acetic acid, 50 mL (v/v) ethanol

dH_2O was added to a final volume of 400 mL.

2.9.4 Buffers and solutions for His-tag purification by immobilized-metal affinity chromatography

Buffer A

20 mM Tris-HCl (pH 7.4), 500 mM NaCl, 20 mM imidazole.

Buffer B

20 mM Tris-HCl (pH 7.4), 500 mM NaCl, 500 mM imidazole.

Tris-buffered saline (TBS)

10 mM Tris-HCl (pH 7.4), 150 mM NaCl.

2.9.5 Buffers for microscopy

Phosphate-buffered saline (PBS)

8 g NaCl, 0.2 g KCl, 1.44 g Na_2HPO_4, 0.24 g KH_2PO_4

The components were dissolved in 800 mL dH_2O and pH was adjusted to 7.4. dH_2O was then added to a final volume of 1 L.

1.2% agarose gel in PBS

0.24 g certified molecular biology agarose, 20 mL PBS

Dissolved by heating in a microwave oven and stored at 55 °C until use.

PBS + Tween 20

0.05% (v/v) Tween-20 in PBS.

2.9.6 Solutions and buffers for competent cell preparation

0.1 M $CaCl_2$

Sterile filtered using a 0.2 µm filter.

15% glycerol with 0.1 M chloride

30% glycerol, 0.2 M calcium chloride dihydrate

Equal parts mixed.

2.9.7 Other solutions

1 M Tris-HCl buffer

121.14 g Tris base

The components were dissolved in 200 mL with dH_2O and adjusted to pH 7.4 by HCl. dH_2O was then added to a final volume of 250 mL.

100 mM isopropyl-β-D-thiogalactoside (IPTG)

Sterile filtered using a 0.2 µm filter.

3. Methods

3.1 Growth and storage of *Escherichia coli* and *Streptococcus dysgalactiae*

Escherichia coli parental strains BL21 and DH5α and *Streptococcus dysgalactiae* parental strains MA201, 015 and 047 were used for the several experiments conducted in this work.

The *E. coli* strains (Table 2.2, section 2.1) were grown under aerobic conditions at 37 ℃, either in a lysogeny broth with shaking (200 rpm) or on lysogeny agar plates prepared with 1.5% (w/v) agar. Both LB medium and LA plates were supplemented with ampicillin (100 μg/mL) for correct selection of transformants. The *S. dysgalactiae* strains (Table 2.3, section 2.1) were either grown in BHI medium or C-medium at 37 ℃ without shaking. Moreover, glycerol stocks of all bacteria were made by adding a final concentration of ~15 % (v/v) glycerol to a log-phase culture. These were then stored at -80 °C. Start cultures of *S. dysgalactiae* were made and stored the same way. For short-term storage, the *E. coli* culture tubes or plates were kept and stored at 4 °C.

3.2 Plasmid isolation

Bacterial plasmids are independently replicating pieces of extrachromosomal circular DNA of varying lengths, often carrying beneficial genes such as antibiotic resistance. These plasmids are taken advantage of in several ways for different purposes as they can be used as tools for cloning and manipulation of genes, and are therefore often experimentally referred to as vectors (106). By transformation, these vectors can then be introduced into bacteria (section 3.7) and copied into larger quantities following bacterial division. For this project, the pRSET A vector was used as a high-level expression vector to overexpress genes (i.e., $cbpD^{SD}$) in *E. coli* BL21, in order to produce recombinant proteins necessary for the present work (section 3.9) (107).

The pRSET A vector was isolated from *E. coli*. DH5α, using the E.Z.N.A.® Plasmid Mini Kit I (Omega Bio-tek) which features a chaotropic salt-based silica column for purification. Following the manufacturers protocol, *E. coli* cells were harvested from 1-5 mL cultures by centrifugation at 4000 x g for 5 minutes. The supernatant was discarded, and the acquired pellet was resuspended in 250 μL Solution I (resuspension buffer), which contains ribonuclease A (RNase A) to eliminate any RNA contamination. The mixture was then transferred to a clean 1.5 mL microcentrifuge tube, followed with further lysis by the addition of 250 μL Solution II (alkaline lysis buffer). After inverting the tube gently and observing a clear lysate, 350 μl solution III (neutralization buffer)

was added to neutralize the acidic pH. This caused precipitation of cellular debris with chromosomal DNA and proteins, while the vector DNA remained in the solution. The precipitated debris was pelleted at 20,000 x g for 10 minutes, and the cleared supernatant was transferred to an equilibrated HiBind® DNA Mini Column (50 µl Equilibrium Buffer). The column contains a silica membrane constituting the DNA binding technology, so that after the addition of 500 µL HBC buffer to the sample, the water molecules capability of keeping the DNA soluble is reduced due to high salt conditions. This allows the negatively charged phosphates in the DNA-backbone to bind to the net positively charged hydrogen atoms of the hydroxyl group in the silica membrane instead. The bound plasmid was then washed with 700 µL DNA Wash Buffer. To remove any residual ethanol left over from the wash buffer, the column was dried by centrifugation at maximum speed for 2 minutes. The column was transferred to a clean 1.5 mL microcentrifuge tube, and the plasmid was eluted using 30-40 µL Elution Buffer (low salt) followed by incubation at room temperature for 1 minute and centrifugation. The concentration of the isolated plasmid was determined using the NanoDrop™ 2000 before being stored at -20 °C (section 3.3).

3.3 Quantification of DNA and proteins

Microvolume quantification of DNA and proteins was done using the NanoDrop™ 2000 (ThermoFisher Scientific™) spectrophotometer. For DNA, the spectrophotometric analysis is based on the principle that nuclei acids absorb ultraviolet (UV) at a wavelength of 260 nanometers (nm), where the amount of light absorbed is proportional to the concentration of DNA in the sample. On the other hand, determination of protein concentration is performed by measuring absorbance at 280 nm, and is based on the absorbance of UV light by the aromatic amino acids tryptophan (W) and tyrosine (Y) as well as disulphide-bonded cysteine residues (108). The ProtParam tool (109) was employed to calculate the theoretical molecular weight (kDa) and the molar extinction coefficient (ε) of various $CbpD^{SD}$ protein sequences, with the latter being the weighted sum of the 280 nm molar absorption coefficients of the three mentioned amino acids (1.1) (110, 111).

$$\varepsilon = (nW \times 5500) + (nY \times 1490) + (nC \times 125), \quad (1.1)$$

The factor n is the number of each type of residue (Trp, Tyr and Cys) per molecule of protein, and the stated values are the amino acids molar absorptivities at 280 nm. In light of this and utilizing Beer's law, absorbance (A_{280}) of a sample is dependent on the molar extinction coefficient (ε), the

molar concentration (c) and the light path length in centimeters (L) which is usually assumed to be equal to one, resulting in the following formula (1.2) (110, 111).

$$A\lambda = \varepsilon cL \quad (1.2)$$

The NanoDrop™ utilizes this formula to calculate the protein concentration (c) in mg/mL when given the molecular mass and extinction coefficient of the protein. Two µL of the sample was pipetted directly onto the pedestal of the instrument, and concentration of either nuclei acids (in ng/mL) or proteins (in mg/mL) of the samples were readily available.

3.4 Polymerase chain reaction

The polymerase chain reaction (PCR) is a revolutionary *in vitro* technique used to amplify a stretch of distinct DNA sequence, where it can amplify infinitesimally small amounts of DNA into quantities amenable for analysis (112). These DNA amplicons are subsequently used in various ways, for instance for sequencing, visualization by gel electrophoresis or for cloning into a plasmid, making it an indispensable and versatile tool in modern molecular biology. The fundamental principle of PCR is the use of a DNA polymerase to amplify the DNA sequence by using a pre-existing DNA molecule (112, 113). The Norwegian researcher Kjell Kleppe demonstrated the principle of PCR in 1969 at a The Gordon Conference and published its practical use in 1971 (114), but in lack of a DNA polymerase surviving the heating steps required to melt DNA in a PCR reaction, he had to add new enzyme for each cycle. The method was further developed in the 1980s by Kary B. Mullis, who instead used a thermostable DNA polymerase acquired from the thermophilic bacterium *Thermus aquaticus*. The DNA strands could therefore be synthesized repeatedly, paving the way for the development of fully automated PCR using thermocyclers (115, 116). The PCR reaction requires other essential components as well: DNA serving as the template, primers and the four deoxynucleotide triphosphates (dNTPs) dATP, dCTP, dGTP and dTTP (117, 118). Primers are short oligonucleotides used to specify the DNA sequence to be amplified and must be sequence complementary to the desired target DNA, where both a forward and reverse primer are necessary for flanking the target region (113).

PCR consist of a three-step cycle of thermal denaturation of the dsDNA, annealing of primers to target sequence and extension performed by the thermostable DNA polymerase from the 3’-OH ends of primers. These steps constitutes only one cycle, but where for each cycle the PCR product

is doubled and is usually repeated 25-30 in order to yield millions copies of identical DNA (118). Primers and the dNTPs will become depleted, as well as that the number of DNA templates exceeds the polymerase supply, causing the DNA synthesis to become less efficient if the cycles are repeated far over 30 cycles (116).

The first step of the PCR cycle involves separating the DNA strands by heating the mixture to 94-98 °C. At this temperature range, the hydrogen bonds between the complementary base pairs of the two DNA strands become separated, causing denaturing of the DNA molecule into single strands (118, 119). During the second step of the cycle, lowering the temperature to between 50-72 °C allows for annealing of the primers to their respective complementary sites on the single stranded DNA molecule (119). Primer design is an important aspect of PCR, where primer length and sequence composition needs to be considered. Primers usually consist of 18 to 22 nucleotides as this lowers the probability of a primer binding to a random sequence. In regard to the primer sequence, the primers melting temperature (T_m) needs to be calculated and refers to when half of the primer is hydrogen bonded to the DNA sequence and the other half is not. This is influenced strongly by both G/C content and length (112). Too low T_m can result in non-specific annealing, and if too high, may result in no annealing of the primers to the template (120). In this work, the primers were designed to have a T_m between 60-65 °C. Finally, in the extension phase, the temperature is raised to 72 °C which is the optimal working temperature for the DNA polymerases used in this work (Phusion® and Red*Taq*®). The polymerase can thereby synthesize a new complementary DNA strand in a 5' $\rightarrow$ 3' direction from the free 3' OH end of the primers. The synthesized extension product produced at the end of this cycle is used as a template for subsequent cycles, where theoretically one template molecule is amplified at the rate of 2^n (n = the number of cycles). This produces the exponentially growing population of identical DNA molecules (118, 119). Figure 3.1 illustrates the three-step cycle of PCR, with a denaturing step, annealing step and finally an extension step where for each step the typical temperature ranges are given. Figures 3.2 and 3.3 displays the specific temperatures and durations applied in this work, respectively for each DNA polymerase used. The PCR conditions are optimized for each DNA template and primer pairs used (116).

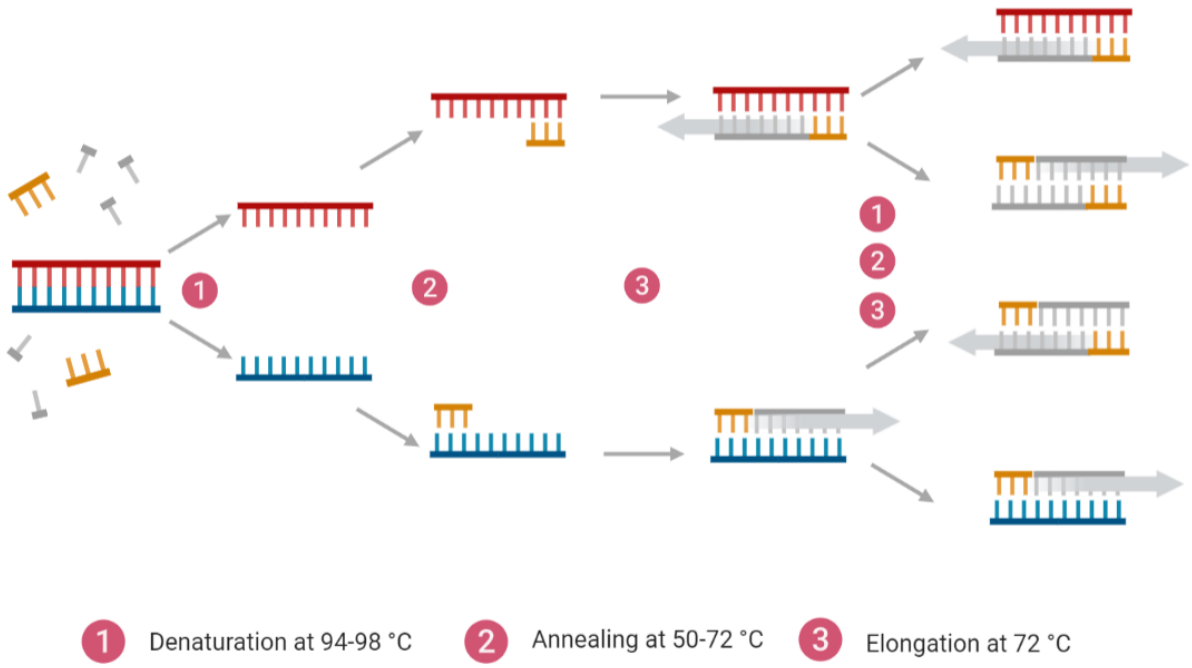

Figure 3.1: A general schematic representation of the three-step cycle of PCR, where the double-stranded DNA is displayed in a red and blue, the primers in orange and the free dNTPs in grey. Denaturation (1) occurs first and cause the DNA molecules to separate into two separate single-stranded DNA molecules. This occurs due to the high temperature disrupting the hydrogen bonds between the complementary base pairs of the opposite strands (94-98 °C). Next is the annealing (2) of the synthetic oligonucleotide primers to the 3' end of the complementary sites of each DNA strands at temperatures 50-72 °C. This allows elongation (3) by the DNA polymerase, where its optimal temperature is usually 72 °C, by adding these free dNTPs to the free 3' OH end of each primer. Theoretically, for each cycle, the number of double-stranded DNA should double. Figure adapted from BioRender (14).

The Phusion® High-Fidelity (HF) DNA polymerase (New England BioLabs®) and the Red*Taq*® DNA polymerase (WVR Life Sciences) were the two DNA polymerases used throughout this work. The Phusion® DNA polymerase was used when precise accuracy of the PCR product was necessary (e.g., sequencing and cloning), as it has a 50X higher fidelity than Red*Taq*® due to its unique structure: a *Pyrococcus*-like enzyme fused with a processivity-enhancing domain. This confers the polymerase with a 3' → 5'exonuclease activity and provides it with speed (30 seconds per kilobase) (121). Red*Taq*®, on the other hand, lacks this proofreading ability and was primarily used for screening colonies when the fidelity was not as important. It also has lower speed, with approximately one min per kilobase (112).

Tables 3.1 and 3.2 displays the various components for PCR either with the Phusion® HF polymerase or the Red*Taq*® polymerase, respectively. The Red*Taq*® polymerase comes as a

ReadyMix™ which contains a reaction buffer with $MgCl_2$, dNTPs, the *Taq* polymerase and a 2x loading dye. Primers, DNA template and dH_2O were the only components needed to be added separately. On the other hand, the Phusion® HF polymerase does not come as a ready master mixture, and a 5x Phusion® HF buffer, dNTPs and Phusion® HF polymerase must be added separately, in addition to the primers, DNA template and dH_2O. A 6x loading dye is also added after PCR amplification, before being loaded onto the agarose gel (section 2.9.2). The appropriate primers are listed in Table 2.4 in section 2.2.

Table 3.1: Components required for the Phusion® High-Fidelity polymerase PCR. The various components are listed on the left-hand side of the table, with the final concentration or volume for the corresponding component found on the right-hand side of the table. The final volume for this PCR-setup was 50 µL.

Components for Phusion® PCR	**Final concentration/ volume**
5x Phusion® High Fidelity Buffer[a]	1x
10 µM Primer F	0.5 µM
10 µM Primer R	0.5 µM
10 mM dNTPs	200 µM
DNA template	< 250 ng
Phusion® HF polymerase	1 unit/50 µL PCR
dH_2O	Adjust to a final volume of 50 µL
Final volume of 50 µL	

a. The 5x Phusion® High Fidelity buffer contains 1.5 mM $MgCl_2$ (121).

Table 3.2: Components required for the Red*Taq*® PCR. The various components are listed on the left-side of the table, with the final concentration or volume for the corresponding component found on the right-side of the table. The final volume for this PCR-setup was 10 µL.

Components for Red*Taq*® PCR	**Final concentration/ volume**
2x Red*Taq*® ReadyMix™ PCR reaction mix[a]	1x
10 µM Primer F	1 µM
10 µM Primer R	1 µM
DNA template	1 µL
dH_2O	Adjust to a final volume of 10 µL
Final volume of 10 µL	

a. Contains a Red*Taq* reaction buffer (Tris-HCl pH 8.5, $(NH_4)_2SO_4$, 3.0 or 4.0 mM MgCl2, 0.2 % Tween®), 0.4 mM of each dNTP, 0.2 units/µL VWR Taq polymerase and 2x red dye (122).

Figure 3.2 displays the thermocycling conditions when the Phusion® HF DNA polymerase was utilized for PCR. The three-step cycle is consistent of what was previously mentioned, with a denaturation step, an annealing step and an elongation step. Additionally, it should also be recognized that an initial denaturation step and a final elongation step also occurs, but only for a

single cycle (x 1). The temperature and duration of each step is displayed, with an indefinite hold at 4 ℃ to allow the preservation of the samples in the end. The conditions displayed here are adjusted after the length of the DNA template and the T_m of the primers, where the annealing temperature is set 2 ℃ below.

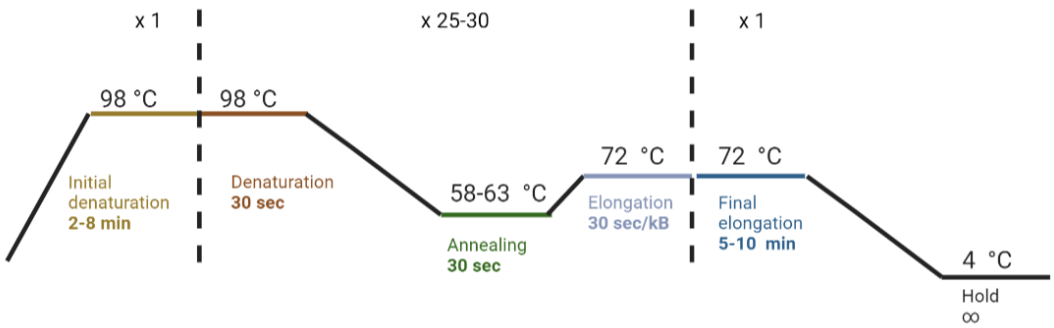

Figure 3.2: The standard program for Phusion® High Fidelity DNA polymerase PCR, with an initial denaturation, denaturation, annealing, elongation, final extension and a hold step. Each step of the program is given a colour which corresponds to the description and duration of the step underneath. Above each step, the temperature is also displayed. The number of cycles are listed on top, as indicated by the x. The conditions are affected by both the length of the DNA template, efficiency of the polymerase, as well as the T_m of the primers.

Figure 3.3 displays the thermocycling conditions when the Red*Taq*® DNA polymerase was utilized for PCR. Similarly for the Phusion® PCR, an additional pre-denaturation and final extension steps are also included and only occurs as a single cycle. The temperature and duration of each step is also displayed, with an indefinite hold for 4 ℃ in the end for preservation. The conditions displayed here are adjusted after the length of the DNA template and the T_m of the primers.

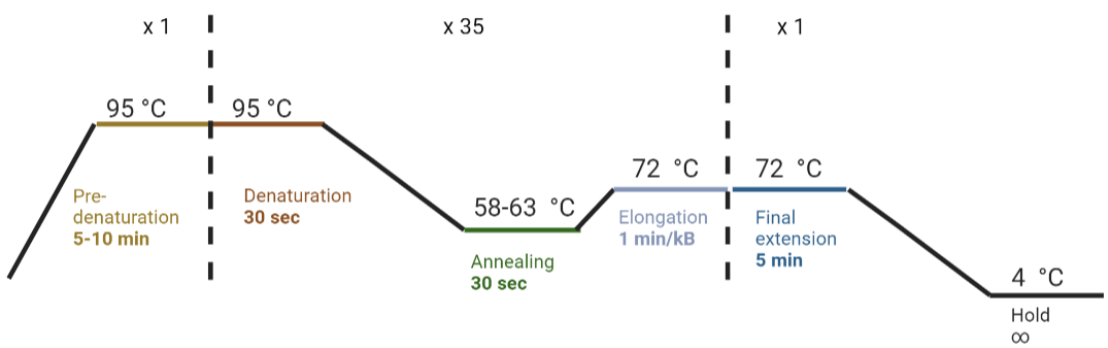

Figure 3.3: The standard program for Red*Taq*® DNA polymerase PCR, with a pre-denaturing step, denaturing step, annealing step, elongation step, final extension step and finally a hold. Each step of the program is given a colour which corresponds to the description and duration of the step underneath. Above each step, the temperature is also displayed and the number of cycles (indicated by the x). The conditions are affected by both the length of the DNA template, efficiency of the polymerase, as well as the T_m of the primers.

3.4.1 Colony PCR

Colony PCR was used as a method of rapid screening of bacterial colonies in order to confirm the presence of correct construct after a transformation into *E. coli* (123) (section 3.7). The primer pair used for colony PCR anneal to sites flanking the insert and gave a PCR product of known size if the desired construct was present (Table 2.4, section 2.2).

After a transformation and visible growth of colonies on the LA plate, a sterile toothpick was used to pick an *E. coli* colony. The toothpick was tapped two times at the bottom of an empty sterile PCR tube to leave a footprint of cells before it was transferred to a culture tube containing appropriate growth medium and ampicillin (100 µg/mL). 10 µL Red*Taq*® PCR mixture was then added to the PCR tube (see Table 3.2), which was placed directly into the PCR thermocycling machine (ProFlex™ PCR system, Applied Biosystems™) with an extended duration of the pre-denaturation step to 10 minutes. It was also possible to lyse the *E. coli* colony in 10 µL dH_2O with microwaving at max for 2 minutes, though this was not done very often. These heating steps were necessary to release the plasmid from the bacterial cells, which was then used as template for PCR. The standard thermocycling conditions for the Red*Taq*® screening can be found in Figure 3.3. Lastly, the PCR products were visualised and verified using agarose gel electrophoresis (section 3.5).

3.4.2 Overlap-extension PCR

Overlap-extension PCR (OE-PCR) is a variation of PCR, which can be used for site-directed mutagenesis of a DNA sequence or fusing DNA fragments not normally next to each other (112, 124). The most crucial part of OE-PCR is the use of overlapping primers, meaning they have a 5' overhang complementary to the end of the DNA fragment to be fused with or also contains a site-directed mutation (125). In this work, OE-PCR was employed in order to introduce site-directed mutations of *cbpD*^{SD} and *sfgfp-cbpD*^{SD} and removal of the DNA sequence encoding the SH3b-domain of sfGFP-CbpD^{SD}. The Phusion® HF DNA polymerase was made use of in amplifying the DNA fragments, with components necessary for the PCR reaction mixture presented in Table 3.1 and the thermocycling conditions presented in Figure 3.2.

OE-PCR is a two-step process (Figure 3.4), where the first step, by two parallel PCR reactions, creates intermediate PCR fragments (AB and CD) having sequence overlaps so that the 3' end of one fragment (AB) overlaps at the 5' end of another fragment (CD). In the second step of OE-PCR,

the fragments are used as template and will hybridize at their overlapping, complementary regions. The overlaps of each strand can thereby serve as "oversized" primers for extension by the Phusion® DNA polymerase, resulting in the hybrid gene product (AD). By including the forward and revers primer flanking PCR products 1 and 2, the fused PCR product can be amplified by conventional PCR. Finally, the product is visualised, verified and separated by gel electrophoresis (section 3.5), and purified using the NucleoSpin® Gel and PCR Clean-up kit (Macherey-Nagel®) (section 3.5.2).

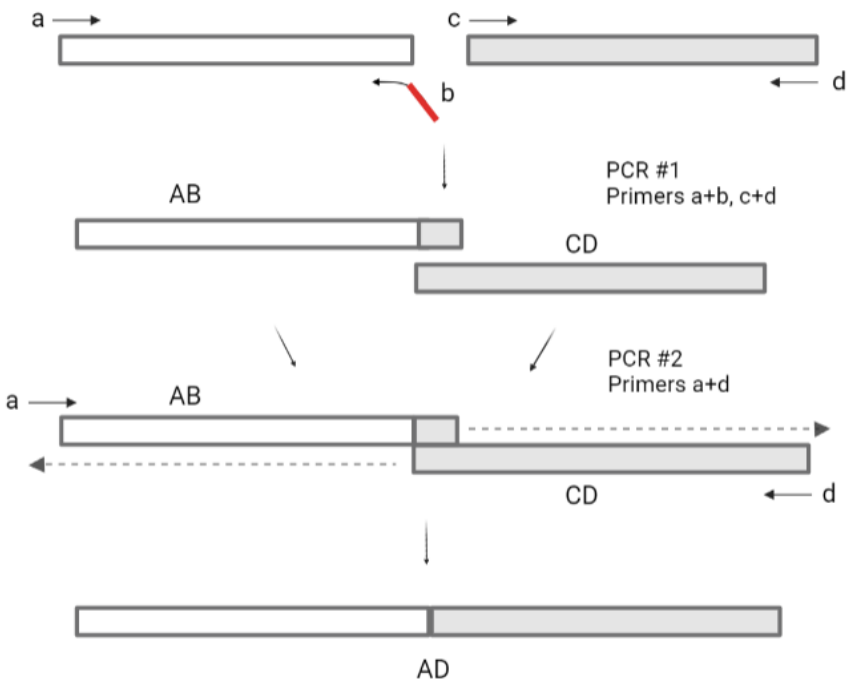

Figure 3.4: Schematic overview of OE-PCR, which is a two-step PCR process. In the first step of PCR (PCR #1), the intermediate PCR products AB and CD are created, where AB fragment has an overlapping region with fragment CD due to the use of overlap primer b (red tail) with 5' overhang complementary to the 3' end at the other fragment. In the next step, AB and CD fragments are used as templates, and hybridize together in order to serve as primers for Phusion® DNA polymerase. Primers a and d are also included to amplify the final hybridization product AD. Created with BioRender (14).

3.4.2.1 SH3b domain deletion

In this work, as already mentioned, OE-PCR was used to remove the SH3b encoding domain positioned between the *sfgfp* and the *cbpD* coding sequences, giving rise to *sfgfp-cbpD*$^{\Delta SH3b}$ (Figure 3.5). The pRSET A vector with the *sfgfp-cbpD*^{SD} coding sequence was used as template, along with flanking primers pRSET(F)+pRSET(R) and internal primers JAA2+JAA1. See section 2.2, Table 2.4 for primer sequences. In the first step of OE-PCR, a reverse primer (JAA2) complementary to sequence *sfgfp* and with a 5' overhang was used in order to create the complementary junction

between *sfgfp* and the conserved domain coding sequence of *cbpD*. This made it possible in the next step of OE-PCR for fusion of the *sfgfp* amplicon and the *cons.dom*$^{CbpD\text{-}SD}$, thereby allowing extension by Phusion® DNA polymerase. Two flanking master primers were also used to mark the 5' ends of all strands and amplify the final hybridization product *sgfp-cbpD*$^{\Delta SH3b}$.

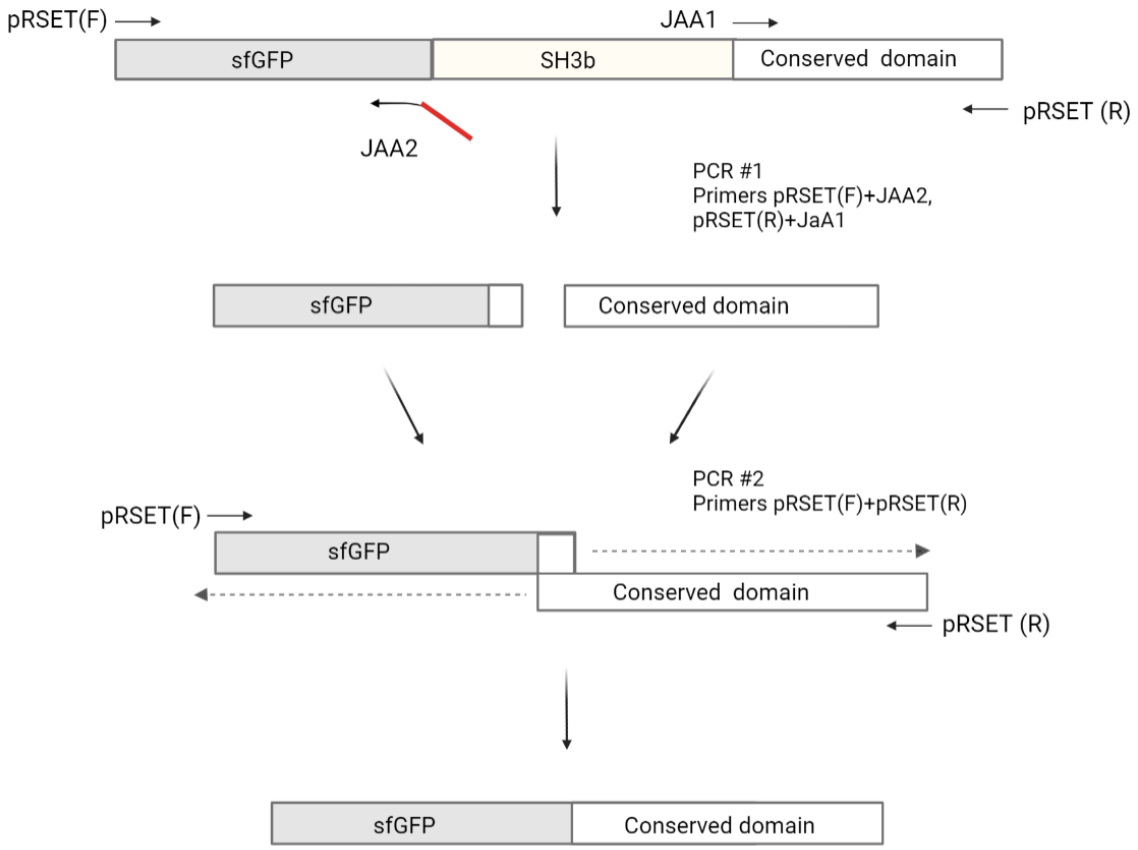

Figure 3.5: Schematic overview of overlap-PCR when used for fusing DNA sequences not normally next to each other, in this case removing the SH3b encoding domain (187 bp). Primers are displayed as arrows, pointing in a 5' → 3' direction and the different colored boxes indicate the DNA sequences of *sfgfp-cbpD* (1371 bp). In the first step of PCR, amplification of the sfGFP coding sequence (744 bp) and amplification of the part in *cbpD* encoding the conserved domain (360 bp) occurs in two parallel reactions. The sfGFP amplicon has a complementary region at the 3' end due to the reverse primer with 24 bp 5' overhang (JAA2, red tail) complementary to the forward primer (JAA1). The two fragments are combined in a single PCR reaction and will thus anneal at their complementary regions. This, in combination with the outside flanking primers pRSET(F) and pRSET(R), creates the hybridization gene product *sfgfp-cbpD*$^{\Delta SH3b}$. Created with BioRender (14).

3.4.2.2 Site-directed mutagenesis of CbpD^{SD} and sfGFP-CbpD^{SD}

OE-PCR was also employed to create two point-mutated versions of CbpD^{SD} (CbpD^{C65A} and CbpD^{G302A,G303A}) and four point-mutated versions of sfGFP-CbpD$^{\Delta SH3b}$ (sfGFP-CbpD$^{\Delta SH3b,L300S,A301S}$, sfGFP-CbpD$^{\Delta SH3b,G302A,G303A}$, sfGFP-CbpD$^{\Delta SH3b,Q329A}$, and sfGFP-

CbpD$^{\Delta SH3b,W365A}$). Site-directed mutagenesis involves introducing nucleotide alterations into the target sequence by changing specific DNA bases, e.g., altering specific codons to change one or more amino acids in a protein sequence. The OE-PCR principle remains the same, although the reverse internal primer with the overhang carries the mutation of interest, in addition to being complementary to the forward primer of the next fragment in order to create these overlapping junctions. The next step of OE-PCR can then occur, with the amplicons of the previous step annealing at their complementary regions and become extended by the Phusion® DNA polymerase. Flanking master primers pRSET(F) and pRSET(R) were yet again utilized to amplify the fused recombinant DNA fragment.

3.5 Gel electrophoresis

Gel electrophoresis is a crucial tool for separating, visualizing and purifying DNA, RNA and proteins by primarily their intrinsic electrical charge and secondarily by their size and shape when applied to a porous gel matrix and subjected to an electrical field (118, 126). This method was developed by Arne Tiselius, who saw that molecules would migrate along an electric field if an electrical charge was laid over a sheet of damp paper with a solution containing some electrically charged molecules (127). It has since, conceivably, become the most widely used physical method in molecular biology. The gel material for electrophoresis in the present work consisted either of agarose or polyacrylamide, depending on the particles to be analyzed. Agarose gel electrophoresis was used to separate and visualize DNA molecules, whilst sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) was used to separate and visualize proteins. A description of SDS-PAGE can be found in section 3.10.

3.5.1 Agarose gel electrophoresis

Agarose gel electrophoresis, as previously mentioned, is used for separating various DNA fragments of different sizes. The agarose used is a linear polymer consisting of repeating units of agarobiose isolated and purified from certain red seaweeds (128, 129). When the agarose is heated up in a buffer and set to cool in a process of gelation, the agarose polymers will associate non-covalently and form a molecular sieve of various sized pores. In order to separate the DNA, the sample(s) is loaded into pre-cast wells found at the top of the gel and an electric current is applied. Since DNA has a uniform mass/charge ratio as the phosphate backbones are negatively charged at physiological pH, the molecules will collectively migrate towards the positively charged anode at

the other end of the gel. The distance the DNA fragment have travelled is inversely proportional to the log of its molecular weight, meaning that smaller DNA fragments travel faster through the gel than larger ones (129). In this research, the fluorescent dye peqGREEN was also used to visualize the DNA after separation and was added to the gel right before casting. When bound to DNA, the dye will fluoresce under UV light (130).

In this work, a 1% (w/v) agarose gel was used, with the recipe found in section 2.9.2. After mixing the agarose with 1x TAE (40 mM Tris-Acetate, 1 mM EDTA), the mixture was brought up to a boil by a microwave with some pauses throughout to swirl and allow the components to mix well. When the agarose had completely dissolved, the flask was cooled to approximately 65 °C. The agarose mixture was not cooled enough to allow for gelation, but rather to avoid bending the gel casting tray. After some swirls under cold water, peqGREEN (1 µL/50 mL) was added into the flask and the solution was poured into a cast. Appropriate combs were also placed in the gel to create the wells. After firming up by allowing it to set at room temperature, the combs were removed and the gel was moved to a gel electrophoresis chamber filled with TAE buffer. The samples were finally applied onto the gel within each well. In this work, only PCR products and restriction cut fragments were separated by gel electrophoresis. For screening of PCR products, where the Red*Taq*® polymerase was used, no loading dye had to be used as it was already present in the PCR mix. PCR products produced by the Phusion® HF polymerase, however, had to have a 6x loading buffer mixed with it, resulting in a final concentration of 1x. The loading buffer contains sucrose which raises the density of the DNA sample and allows it to sink to the bottom of the well, as well as bromophenol blue to make the sample visible under application to the gel (section 2.9.2). A 1 kb DNA ladder with DNA fragments of known sizes (0.5 kB to 10 kB) was also applied to one of the wells (5 µL), in order to estimate the sizes of the DNA fragments of the sample. The electrophoresis was performed at 90 V for 20-25 minutes in a Mini-Sub Cell GT Cell (Bio-Rad) to allow the DNA fragments to separate. Lastly, the fragments were visualised under UV light in the GelDoc-1000 (Bio-Rad).

3.5.2 PCR clean-up and extraction from gel

PCR products and other DNA fragments were cleaned free from primers, dNTPs, polymerase enzymes, salts or restriction enzymes, in order to be used for further work (i.e., restriction digestion,

ligation, sequencing or OE-PCR). This was done using the Nucleospin® Gel and PCR Clean-up kit by following its user manual.

If the whole PCR product was loaded onto the gel, the correct DNA band was excised from the agarose gel visualised by UV light, by using a clean scalpel. The gel piece was then dissolved in NTI buffer (~200 µL buffer per 100 mg agarose gel) at 55 °C for 5-10 minutes and transferred to a Nucleospin® Gel and PCR Clean-up Column. For screening, only a small amount of the PCR was necessary to load into the gel. The NTI buffer was therefore directly added to the PCR sample (2-3x the PCR sample volume), and then transferred to the Nucleospin® column. The column was subsequently centrifuged at 11,000 x g for 30 seconds, allowing the DNA to bind to the silica column due to the presence of chaotropic salts in the NTI buffer (as described in section 3.2). The silica membrane was washed with 700 µL buffer NT3 by centrifugation to remove any contaminations, followed with yet another centrifugation to remove any residual ethanol. Lastly, the Nucleospin® column was placed in a clean 1.5 mL Eppendorf tube and the DNA was eluted with 15-30 µL of the slightly alkaline buffer NE (5 mM Tris-HCl, pH 8.5). The column was incubated at room temperature for one minute and then centrifuged. The Nanodrop™ 2000 was used to measure the concentration of the eluted DNA (section 3.3), and then stored at -20 °C until further use.

3.6 Restriction digestion and ligation

In DNA cloning, restriction digestion and ligation are widely used to insert a piece of DNA into a vector that can then be transformed into an appropriate bacterium to produce several copies. This is possible as the DNA fragment to be inserted (insert) and the receiving vector (backbone) has to be cut with compatible restriction enzymes in two separate reactions (Figure 3.6). Restriction enzymes are proteins produced from different bacterial species to protect them from DNA deriving from other organisms or phages by cutting their sequence (118). These enzymes can be isolated from the bacterial cells to be used as an indispensable tool in the laboratory for recombinant DNA technology by taking advantage of their specific recognition sites. Several different restriction enzymes have been identified, where each recognizes a specific non-methylated palindromic sequence of either four, six or eight DNA bases (112). When the restriction enzyme recognizes a sequence, it cuts the DNA by catalyzing hydrolysis of covalent phosphodiester bonds between adjacent nucleotides, producing patterns of either blunt or sticky ends. Sticky ends means that the

ends of the dsDNA molecule have unpaired single-stranded overhangs, which can usually reattach to other ends by forming hydrogen-bonded base pairs with other complementary sticky ends of DNA molecules cut by the same enzymes. In principle, this means that the DNA fragment to be inserted and the vector has to be cut with the same restriction enzyme to generate compatible sticky ends. However, these associations are only temporary but can be made permanent when the enzyme DNA ligase catalyzes the formation of covalent bonds that close up the sugar-phosphate backbones of adjacent DNA strands. This produces the recombinant vector (112, 118).

In this work, the restriction enzymes HindIII (5' A$^{\downarrow}$AGCTT 3', 4 bp overhang) and NdeI (5' CA$^{\downarrow}$TATG 3', 2 bp overhang) (New England BioLabs®) were used for cutting the *cbpD*^{SD} and *sfgfp-cbpD*^{SD} fragments and the pRSET A vector. They both belong to the Type II restriction enzymes, meaning that they cut the DNA in the middle of its recognition site and produce sticky ends (112). In addition, the NdeI restriction site includes the ATG start codon for protein translation and therefore cause ligation of the start of the gene, which is useful for expression of the constructs produced throughout this work. A standard restriction digestion reaction for this work is found in Table 3.3. The DNA (insert or backbone) to be digested was mixed with NEBuffer™ 2.1 (New England Biolabs®), the restriction enzymes HindIII and NdeI and dH_2O. The reaction was then incubated at optimal temperature for 1-2 hours, which in this work was in a 37 °C water bath for 1 hour and 30 min. In addition, for the reaction containing the destination vector, the component CIP (Alkaline Phosphatase, Calf Intestinal) had to be added into the mixture after 1 hour into the incubation and with 30 minutes remaining. CIP dephosphorylates the 5' ends of DNA phosphomonoesters, thereby preventing the possibility of self-ligation of the vector. In order to confirm correct restriction cutting, it was verified by gel electrophoresis, cut directly from the gel and purified with the NucleoSpin® Gel and PCR Clean-up kit (section 3.5.2).

Table 3.3: The components necessary for a standard restriction digestion in this work, with final concentration/volume also included. Only restriction enzymes HindIII and NdeI were used.

Components	Final concentration/volume
DNA	1 μg
Restriction enzyme HindIII	1 μL
Restriction enzyme NdeI	1 μL
10x restriction buffer[a]	1x
Alkaline Phosphate, Calf Intestinal (CIP)[b]	1 μL
dH_2O	X μL
Final volume of 40 μL	

a. 1x NEBuffer™ 2.1 contains 50 mM NaCl, 10 mM Tris-HCl, 10 mM $MgCl_2$ and 100 μg/mL BSA, pH 7.9. Supplied as a 10x concentrated stock (131)
b. CIP was added as a component when restriction digesting the pRSET A vector

For ligation, the purified restriction cut DNA fragment and vector backbone were mixed together, along with a 2x Quick ligase reaction buffer (New England Biolabs®), Quick Ligase™ (New England Biolabs®) and dH_2O. An excess of the DNA insert (3:1 in molar) is necessary to ensure proper incorporation of the insert into the vector. The ligation was brought by the Quick Ligase™ which catalyzes the formation of covalent phosphodiester bonds between adjacent vector and insert. The ligase mixture was then incubated for 5 minutes at room temperature, and either stored indefinitely at -20 ºC or chilled on ice before being transformed directly into competent *E. coli* DH5α cells (section 3.7). Table 3.4 lists all of the mentioned components for the ligation reaction, along with their final concentrations or volumes.

Table 3.4: The components necessary for a standard ligation reaction in this work, with final concentration/volume also included. Only the Quick ligase™ was used.

Components	Final concentration/volume
Restriction cut DNA fragment (insert)	6 μL
Restriction cut vector (backbone)	2 μL
Quick Ligase Reaction Buffer (2x)[a]	1x
Quick ligase™	1 μL
dH_2O	X μL
Final volume of 20 μL	

a. 1X NEBNext Quick Ligation Reaction Buffer contains 66 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM DTT, 1 mM ATP, 6% polyethylene glycol, pH 7.6. Supplied as a 2x concentration stock (132).

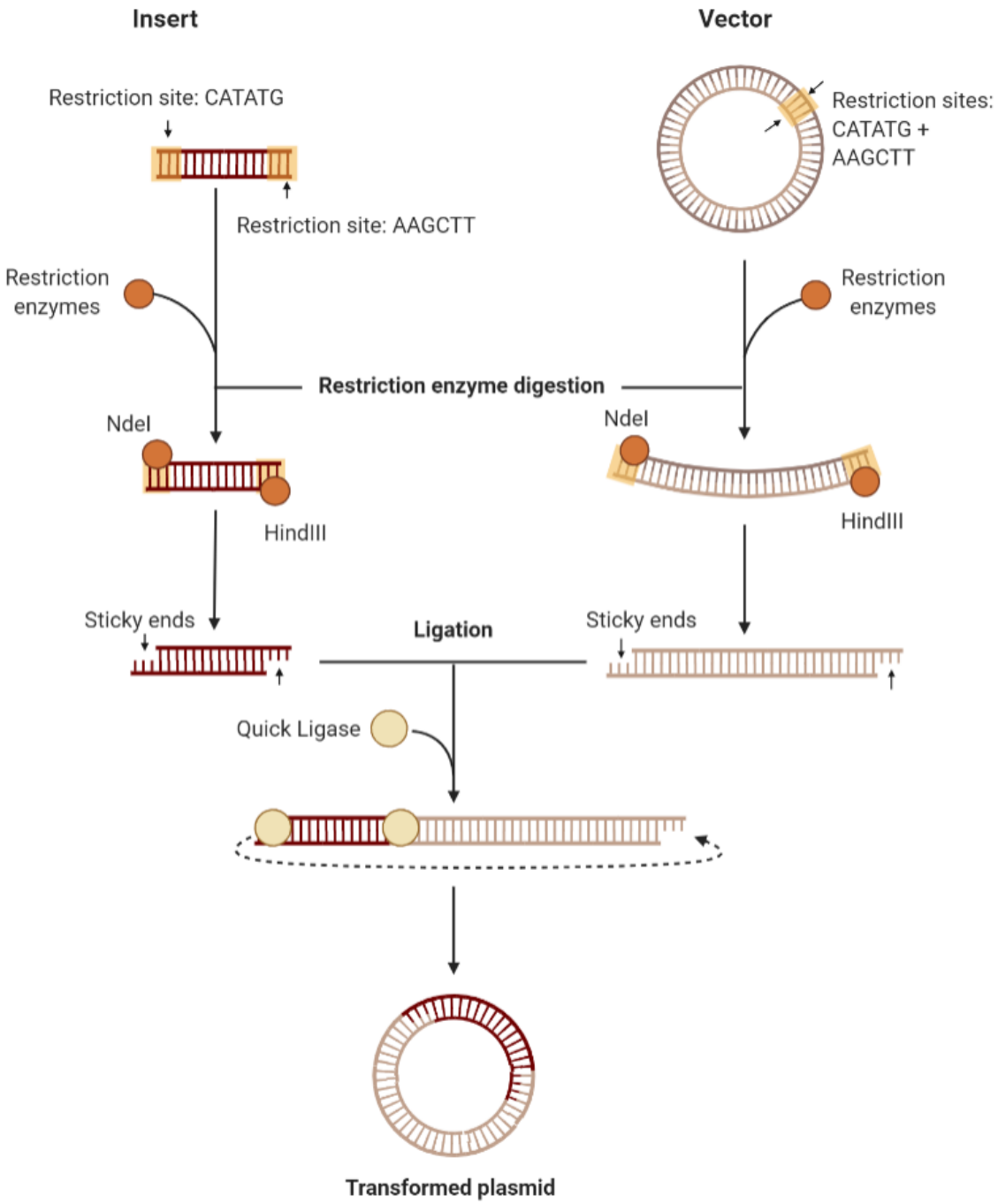

Figure 3.6: Overview of the various steps for restriction digestion and ligation to construct recombinant vector. The DNA insert and vector contain the same restriction sites (CATATG for NdeI + AAGCTT for HindIII) and are thereby cut with the same restriction enzymes NdeI and HindIII. NdeI leaves a 2 bp single-stranded overhang after cutting, while HindIII leaves a 4 bp overhang. This produces compatible sticky ends and cause the insert and vector backbone to be complimentary towards each other. In order to close the gaps together permanently by covalent bonds, ligation by Quick ligase™ occurs. The ligase joins the sugar-phosphate backbone permanently together, and the recombinant vector is made. Created with BioRender (14).

3.7 Transformation of chemically competent *Escherichia coli*

The pRSET A vector was introduced into chemically competent *E. coli* using heat-shock transformation, and as *E. coli* are not naturally competent for genetic transformation, competence had to be developed by a $CaCl_2$-treatment (133). It is believed that ice-cold $CaCl_2$ facilitates introduction of a vector to a cell, although the mechanism behind this remains unknown. Though, it is hypothesized that the divalent calcium ions shield the repulsive forces between the negatively charged phosphate groups on the DNA and the negative charges of the outer membrane of *E. coli*. Cell-membrane disintegration has also been shown to occur during treatment, which increases membrane permeability and makes it accessible for DNA uptake (133, 134).

The following protocol was used in order to make chemically competent *E. coli* DH5α and BL21 cells: 5 mL overnight culture of the strain was grown in LB medium, incubated at 37 °C with shaking. The culture was then diluted to OD_{600} = 0.05 and placed in an Erlenmeyer flask with LB medium (5-10x the previous medium volume) to allow growth to early log phase (OD_{600} = 0.2 – 0.4). The cells were placed on ice for 20 minutes, before being collected by centrifugation at 5000 x g for 5 minutes at 4 °C. The supernatant was discarded, and all subsequent steps were carried out at 4 °C. The cells were resuspended in ½ culture volume in 0.1 M sterile ice-cold $CaCl_2$ and kept on ice for 2 hours. This was followed by centrifugation at 5000 x g for 5 minutes and resuspension in 1/10 culture volume 0.1 M $CaCl_2$ with 15% (v/v) glycerol (section 2.9.6). The now competent *E. coli* cells were stored at -80 °C.

For heat-shock transformation, 50 µL competent *E. coli* DH5α or BL21 cells were thawed on ice and mixed with 1-10 µL of the ligated construct (section 3.6). The mixture was then incubated on ice for 30 minutes, following by a heat shock at 42 °C for 30 seconds. The cells were then cooled on ice for about 1 minute. 450 µL SOC medium (section 2.9.1) was added to the Eppendorf tube containing the cells and then incubated at 37°C with shaking for 1 hour. SOC medium is a nutrient-rich microbial broth and cell recovery medium, and was therefore used to help the *E. coli* cells recover optimally from the stress of heat-shock and pore formation by providing them with a source of carbon, nitrogen and other minerals (135). Finally, the whole bacterial culture was plated on LA plates containing ampicillin (100 µg/mL) for selection of transformants. Colony PCR (section 3.4.1), as well as targeted gene sequencing of the plasmid insert (section 3.8) were used to verify the transformants.

3.8 DNA sequencing

Sanger sequencing is a method for determining the nucleotide order of a sequence and was used to verify correct plasmid construct, as well of those with mutations and deletion of DNA sequences created by OE-PCR (section 3.4.2). The sequencing itself was performed by GATC, Eurofins Genomics (136).

Sanger sequencing was first developed in 1977 by Frederick Sanger (137) and is often referred to as a chain termination method (138). The technique requires a DNA polymerase, primers flanking the sequence area of interest, and a supply of deoxynucleotide and dideoxynucleotide triphosphates. Similar to PCR, there are also several cycles of denaturing, primer annealing and elongation. However, the DNA molecule is mainly determined by dideoxynucleoside triphosphates (ddNTPs: ddATP, ddTTP, ddCTP, ddGTP) carrying a fluorescent dye for each base. These lack the 3'-OH group, and DNA extension comes to a halt as no nucleotide can be added next by the DNA polymerase. The sequencing reaction therefore produces several truncated DNA fragments as the enzyme could either incorporate dNTPs or ddNTPs causing chain termination (106, 112, 138). The Sanger sequence method also employs a capillary system with a polymer matrix for fragment separation in a manner similar to electrophoresis. A laser excites the fluorescent tag found on the ddNTPs as the fragments pass through the system and either a G, A, T or C is recorded (112). A computational software is employed and records the color produced from each fragment which can be displayed as a four-colored chromatogram. Each peak of the chromatogram corresponds to different bases, with often the decoded sequence along the top (106, 112)

The gene to be sequenced was first amplified by Phusion® and verified by gel electrophoresis. The sequencing reaction was set up as follows: 4 µL PCR amplicon (80-100 ng/µL), 1 µL sequencing primers (10 µM) (Table 2.4, section 2.2), and 5 µL dH_2O were all mixed together in separate tubes for each DNA strand. These were sent in for sequences, where a successful read gave about 1000 bases. The sequences were then analyzed using Benchling (139) and nucleotide BLAST by aligning query and subject sequences against each other (140).

3.9 Overexpression of recombinant proteins in *Escherichia coli* strain BL21

CbpD^{SD} and the point-mutated version CbpD^{C65A}, CbpD^{G302A,G303A}, sfGFP-CbpD$^{\Delta SH3b}$, sfGFP-CbpD$^{\Delta SH3b,L300S,A301S}$, sfGFP-CbpD$^{\Delta SH3b,G302A,G303A}$, sfGFP-CbpD$^{\Delta SH3b,Q329A}$ and sfGFP-CbpD$^{\Delta SH3b,W365A}$ were overexpressed in *E. coli* BL21. This strain was used as an expression host as it lacks both a lon protease as well as the ompT outer membrane protease that potentially could degrade the proteins (141). In addition, it expresses the T7 RNA polymerase which is required for transcription of recombinant genes placed behind the T7lac promoter found on the pRSET A vector.

In order to overexpress these recombinant proteins, a 400 mL culture of LB medium with ampicillin (100 µg/mL) and BL21 cells were grown to an OD_{600} of 0.4. IPTG was added to a final concentration of 0.1 mM in order to induce the expression of the recombinant gene. The culture was then placed in a 25 ℃ incubator with 150 rpm shaking for 4 hours. Lastly, the pellet was harvested by centrifuging for five minutes at 7000 x g and 4 ℃ and by decanting the supernatant. The pellet was then stored at -20 ℃ until the following day.

3.9.1 Plasmid for expression by T7 RNA polymerase

In order for high-level protein expression and purification, the pRSET A vector was utilized. This vector is based on the pET (*p*lasmid for *e*xpression by *T*7 RNA polymerase) series of vectors (142). Figure 3.7A illustrates the general features of the pRSET vector, where it contains an ampicillin resistance gene, which allows for correct selection, a pUC origin for high copy replication and growth and a f1 origin for ssDNA production by rolling circle replication. Furthermore, the vector also contains a ribosomal binding site (RBS) for translation of desired gene and an initiation ATG as a translational initiation site for the fusion protein. An N-terminal 6xHis-tag (polyhistidine-tag) encoding sequence which allows for purification on IMAC (section 3.9.3). Lastly, the vector contains a multiple cloning site (MCS) where the desired DNA sequence was inserted. The RBS, ATG initiation site, 6xHis and the MCS are all under the control of a hybrid *T7lac* promoter (107). It is recognized by T7 RNA polymerase, but also contains the binding site for the LacI repressor which blocks transcription. LacI releases DNA upon binding to its allolactose ligand, allowing transcription to occur. Instead of allolactose, which will be metabolized, the non-metabolizable analogue molecule IPTG was used to induce the T7lac promoter (Figure 3.7B.)

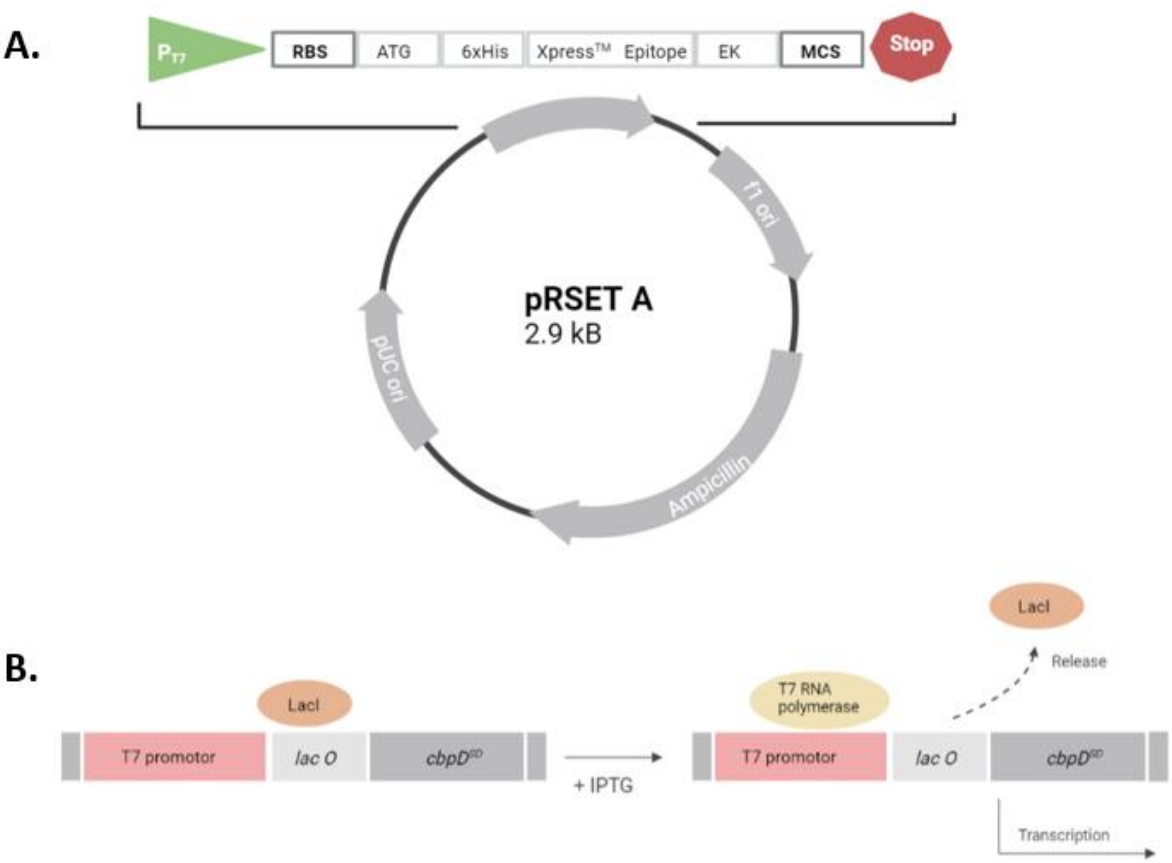

Figure 3.7: A. General features of the pRSET A vector. It contains an ampicillin resistance gene which allows for correct selection of the strain, a pUC origin for high copy replication and a f1 origin for single strand DNA production. A ribosomal binding site (RBS), initiation ATG for translation of recombinant proteins and a N-terminal 6xHis tag is found under the control of a T7 promoter. **B**. Graphical overview of the *t7lac* promoter with a T7 promoter and a *lacO* operator. The desired gene, for instance *cbpD*^{SD}, is found downstream of the operator. When no iso-propyl-thiogalactoside (IPTG) inducer is present, the repressor protein LacI binds to the operator thereby preventing the T7 RNA polymerase (already present in *E. coli* BL21) to express the gene due to blockage. When IPTG is present, it causes LacI to lose affinity and is released from the operator. The polymerase binds to the promoter, thus starting transcription of the gene. Figures created with BioRender (14).

Figure 3.8 represents both the pRSET-His-CbpD^{SD} and pRSET-His-sfGFP-CbpD^{SD} vectors in which the *his-cbpD*^{SD} and *his-sfgfp-cbpD*^{SD} gene has been cloned downstream of the T7lac promoter using the NdeI and HindIII restriction sites. Various versions of CbpD^{SD} and sfGFP-CbpD^{SD} were cloned into the pRSET A vector in a similar fashion. A TEV-proteolytic sequence (N'-ENLYFQG-C') was introduced between the His-tag and *cbpD*^{SD} gene to facilitate removal of the His-tag. However, it was found that this polyhistidine tag did not significantly interfere with the overall function of the CbpD^{SD}, hence, all versions of *sfgfp-cbpD*^{SD} cloned in this work were constructed without the TEV-site (143).

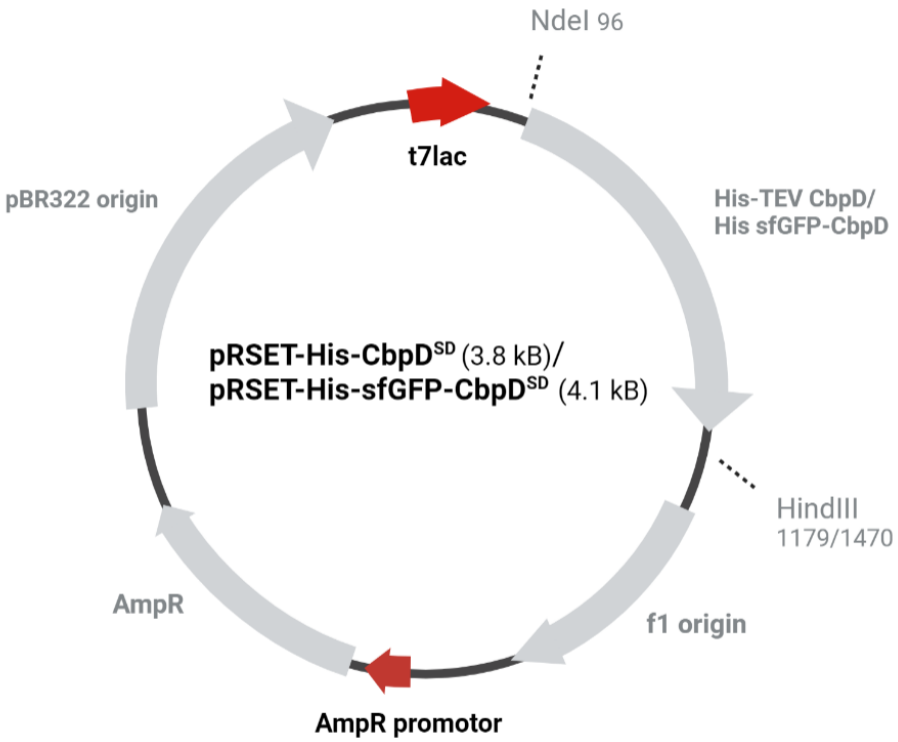

Figure 3.8: Features of the pRSET-His-TEV-CbpD^{SD} and pRSET-His-sfGFP-CbpD^{SD} vectors sharing almost identical characteristics and is therefore displayed as a single vector in this figure. Both vectors were used throughout this work. It contains an f1 origin for rolling circle replication, and *ampR* gene for selection of correct bacteria by ampicillin. In addition, it contains a pBR322 origin for high copy numbers. The desired DNA sequence is cloned into the vector behind a t7lac promoter. Localisation of the restriction sites for NdeI (96 bp) and HindIII (1179 bp for CbpD^{SD} and 1470 bp for sfGFP-CbpD^{SD}) are also included in the figure. It should be taken into consideration that the size of the vector and restriction are dependent on the recombinant gene used. Created with BioRender (14).

3.9.2 Lysis of *Escherichia coli* strain BL21 cells with expressed recombinant proteins

Mechanical lysis of the *E. coli* BL21 cells were performed in order to release the various overexpressed recombinant CbpD^{SD} and sfGFP-CbpD^{SD} proteins, which was done using the High Pressure Cell Disrupter (Stansted®). The pellet, which was obtained as described in section 3.9, was dissolved in 5 mL buffer A as this was the maximum the machinery could handle. Before loading the sample into the Cell Disrupter, 3 rounds of buffer A were loaded in order to flush the system with the right buffer conditions. A 5 mL cell suspension was then loaded, and the cells were subjected to 1 psi pressure before being subjected to ~234 psi pressure leading to cell lysis. The lysate was then centrifuged at 13,000 x g for 5 minutes at a temperature of 4 ℃. The supernatant was transferred to a new tube with the additional use of a sterile filter (0.45 µM pore size) to clear any insoluble material.

3.9.3 His-tag purification by immobilized-metal affinity chromatography

Six tandem histidine residues were fused to the various recombinant $CbpD^{SD}$ and sfGFP-$CbpD^{SD}$ proteins to allow for purification of the proteins. Histidine has a strong affinity towards divalent nickel(II) ions (Ni^{2+}), and the IMAC column HisTrap High Performance (HP) (Cytiva) could therefore be used as it contains cross-linked agarose beads coupled to a chelating group decorated with Ni^{2+} (Ni Sepharose HP affinity resin). The lysate (section 3.9.2) is passed through the column, and the 6xHis-tagged proteins are retained as they are bound tightly to the nickel resin while all other proteins originating from the bacterial host pass through (143). The histidine has a positively charged imidazole functional group where its nitrogen atoms allows for electron donation causing formation of coordinate bonds with the immobilized Ni^{2+} ions (144). Figure 3.9 displays the Ni^{2+}-based IMAC for the purification of proteins with a 6xHis-tag.

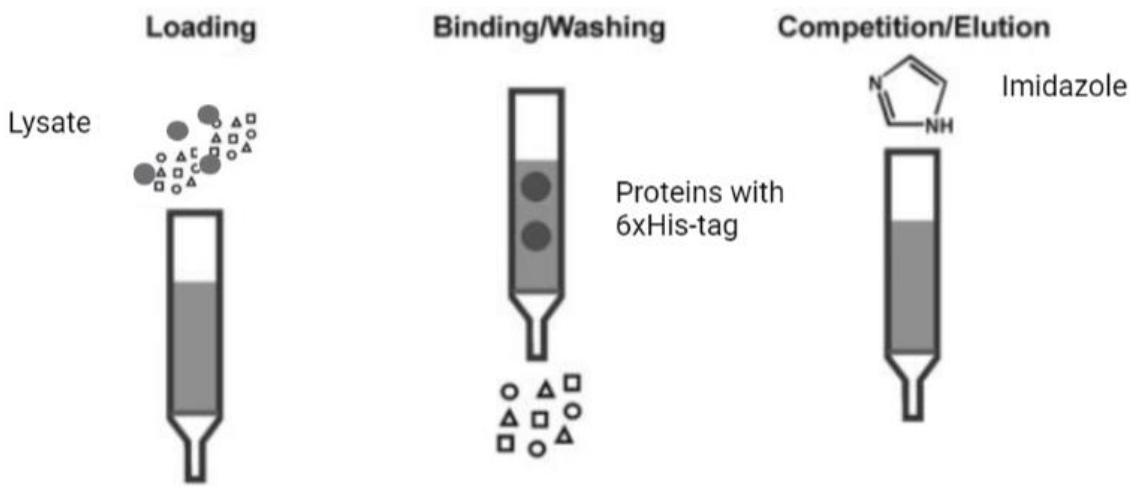

Figure 3.9: Overview of 6xHis-tagged proteins binding to Ni^{2+} ions in an immobilized-metal affinity chromatography (IMAC) column. The lysate mixture is loaded into the column (Loading) where the polyhistidine-tag has strong interactions with the immobilized nickel decorating the chelating group coupled to agarose beads in the column (Binding/Washing). The use of free imidazole cause competition with the 6xHis-tag proteins of the nickel ions, thus displacing the proteins (Competition/Elution). Figure adapted from (145) with BioRender (14).

The clarified cell lysate was manually pushed through a 1 mL HisTrap HP column preequilibrated in buffer A (20 mM free imidazole, see section 2.9.4) using a syringe. Then an ÄKTA pure 25L (Cytiva) instrument was used for column wash and protein elution, where 10x column volume of buffer A was sent through the protein purification system first in order to displace non-specific proteins and other contaminants. Buffer B with 500 mM free imidazole was then sent through in order to elute the 6xHis-tagged proteins as the higher concentration of imidazole competes for the

Ni^{2+}-ions. The flow rate was set at 1 mL/min, with a linear gradient elution of 0-30 mL (0-100%), and the proteins detected by UV light at 280 nm. Finally, the eluted proteins were collected as 1 mL fractions, with the correct fractions followed up with dialysis and SDS-PAGE (section 3.10) to check purification.

3.9.3.1 His-tag protein purification kit on mutated recombinant proteins

An alternative protein purification method was also conducted in correspondence with the amount of recombinant proteins to be purified: a Protino Ni-TED 150 kit for His-tag protein purification in Mini format (Macherey-Nagel®) was used for the various point-mutated sfGFP-CbpD$^{\Delta SH3b}$ proteins as it allowed the purification process to be run in parallel. This IMAC column has a dry silica-based resin precharged with Ni^{2+} ions, where its chelating group (tris-carboxymethyl ethylene diamine based, TED) occupies five of six binding sites in the coordination sphere of the nickel ion. This leaves a remaining site of the Ni^{2+} ion, which will bind to the 6xHis-tagged proteins (146). This method was performed instead of using the Stansted® High Pressure Cell Disrupter and IMAC in combination with ÄKTA pure 25L. The pellet (section 3.9) was dissolved in 2 mL TBS and the cells were mechanical crushed with the use of the bead beating grind and lysis system (FastPrep 24®) with a speed of 6.5 m/s for 2x20 seconds. The Protino Ni-TED kit was then applied on the lysed samples by following the manufacturers protocol (146). The elution gave three fractions with a volume of 240 µL each.

3.9.4 Dialysis

Dialysis of the recombinant proteins was done to remove any imidazole and excess amounts of salts after purification by IMAC (HisTrap and Protino Ni-TED). This was accomplished using either a 3 mL 2K molecular weight cut-off Slide-A-Lyzer™ Dialysis Cassette G2 (ThermoFisher Scientific™) or a molecularporous membrane with cut-off 12-14,000 Dalton (Da) (Spectra/Por®). The preparation and the method for the cassette and the dialysis tubes remained the same, with 10 mM Tris-HCl and 150 mM NaCl mixed together in a 1-liter flask and rehydrating of the membranes by placing either the cassette or tubes in the flask for 1 minute. The sample was then loaded into either the cassette or the tubes and placed in the flask again for 1 hour on a magnetic stirrer. A syringe was then used to recover the sample from the sample ports found on the cassette or carefully inserted from the side of the dialysis tubes. The dialyzed recombinant proteins were quantifiably measured (section 3.3) and finally mixed with 20% glycerol and stored in -80 °C.

3.10 Sodium dodecyl sulfate-polyacrylamide gel electrophoresis

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was used throughout this work to obtain an analytical separation of the recombinant proteins based on their molecular weights. The molecular weight can be estimated by including a marker of known size, and indicates therefore the correct presence of the protein as well as the purity in a given sample. Similarly to agarose gel electrophoresis, the proteins are separated by an electrical field. The gel used, however, is the polymer polyacrylamide, as mentioned in section 3.5. Polymerization of acrylamide monomers and the cross-linking agent `bis`-acrylamide produces a cross-linked matrix acting as a sieve, allowing smaller molecules to migrate faster than larger ones (147, 148) (section 2.9.3). Additionally, the pores of the matrix in which the proteins migrate through can be altered depending on the acrylamide concentration, so the higher the concentration, the smaller the pore size (147). Polymerization is also only initiated by ammonium persulfate and *N,N,N',N'*-tetramethylenediamine, and is an example of a free-radical catalysis reaction. TEMED catalyzes the decomposition of the persulfate-ion of the APS in order to produce a free radical, i.e., a molecule with an unpaired electron. This free radical binds to the acrylamide monomer molecule, which produces a complex with an unpaired electron, and yet another acrylamide monomer molecule can bind. Long linear chains of acrylamide are therefore made, with occasional cross-linking by `bis`-acrylamide (148).

Unlike nucleic acids, which are all negatively charged, the amino acids of proteins might be either positive, negative or neutral. In order to separate the proteins based on their molecular weight, boiling them in anionic sodium dodecyl sulfate and β-mercaptoethanol/dithiothreitol is necessary (112, 148, 149). β-mercaptoethanol/DTT disrupt any disulfide bridges that help maintain the protein tertiary structure, while the amphipathic structured SDS denatures the protein by destroying their 3D structure. In addition, the SDS has a long hydrophobic hydrocarbon tail with a negative charge at the end, which wraps around the backbone of the protein giving it an overall negative charge and prevents the protein from refolding. The hydrophilic sulfate part of the SDS keeps the protein soluble in water and also maintains the proteins' unfolded shape. In this manner, the proteins will migrate towards the positive electrode in the electrophoretic field in a measure proportional to their weight (112, 150).

In this present work, SDS-PAGE was done as following: the protein samples were prepared by mixing 2x SDS loading buffer (section 2.9.3) to a total volume of 20 µL. The loading buffer also contains bromophenol blue for easier application of the sample onto the gel and for monitoring the electrophoresis run, as well as glycerol for solution density. The mixture was then incubated and let boil for 5-10 minutes at 95 °C. In the meantime, the gel casting molds were assembled with a short plate and a spacer plate, as provided by the Mini-PROTEAN® Tetra Handcast Systems (Bio-Rad), and placed on the casting stands. Two different gels were made, a separation gel and a stacking gel. The short stacking gel was placed on top of the separating gel in order to concentrate the protein sample into a tight and sharp band before it entered the main separating gel. This helped to improve the resolution of electrophoresis, as well as allowing for entrance of the proteins at the same time (148). A 12% separation gel and a 4% stacking gel were made as described in section 2.9.3, with freshly made 10% (w/v) APS and 0.005 mL TEMED added last in order to initiate the polymerization. Bromophenol blue dye (0.05 % w/v) was also added to the stacking gel for better visualization of the wells. 3.2 mL of the separation gel was then quickly pipetted between the two plates, immediately followed by adding approximately 3.2 mL of the stacking gel. The correct comb style was then placed within the plates, and the gels were let polymerize for 30 minutes. After polymerization, the gel casting molds were placed in a multi-casting chamber filled with 1x SDS running buffer. The combs were removed, and the samples were added into wells. Electrophoresis was then started using 120 V until separation of the ladder and then 200 V until the bromophenol dye had reached the end of the separation gel. The separated proteins on the gel were removed between the glass plates, and shaken in Coomassie Brilliant Blue for 10 minutes to stain the proteins (section 2.9.3). A destainer solution was then added onto the gel in order to remove any unbound background dye, and was let sit for 1 hour (section 2.9.3). Destainer was subsequently removed, and water was added to allow for some swelling of the gel and an image was taken with the Azure Imager c400 (Azure Biosystems).

3.11 Isolation of bacterial cell wall

In order to determine the binding pattern of sfGFP-CbpD^{SD} against cell wall with and without teichoic acids, isolating the sacculi was necessary. The following protocol is a hybrid of protocols obtained from Eldholm et al. (73) and Vollmer (151).

A 100 mL *S. dysgalactiae* strain MA201 culture grown in BHI medium to an OD_{550} = 0.35 was harvested at 4000 x g for 10 minutes. The cellular pellet was dissolved in 5 mL 5% (w/v) SDS and incubated at 85 ℃ for 15 min to dissolve the cellular membrane and proteins. The cells were then harvested at 18,000 x g for 15 minutes, and stepwise washed twice with 5 mL 2 M NaCl and then washed four times with 10 mL H_2O. These wash steps were necessary to remove proteins, electrostatically binding to the sacculi, and SDS. The cells were then resuspended in 5 mL 50 mM Tris-HCl (pH 7.4), 20 mM $MgSO_4$ and DNase I and RNase to a final concentration of 10 μg/mL and 50 μg/mL, respectively. These nucleases were added in order for enzymatic breakdown of any DNA and RNA. The mixture was incubated for 1 hour at 37 ℃ with shaking. Finally, $CaCl_2$ was added to a final concentration of 10 mM and trypsin to a final concentration of 100 μg/mL. Trypsin is a protease and therefore cleaves proteins, while $CaCl_2$ was utilized in order to stabilize the enzyme (152). The mixture was incubated overnight at 37 ℃ with shaking. The following day, 8% (w/v) SDS to a final concentration of 1% was added to the overnight mixture, and then incubated at 85 ℃ for 15 min. SDS was used here to dissolve the remnants of the enzymes DNase, RNase and trypsin (153). The cells were harvested by 18,000 x g for 15 minutes, washed stepwise once with 10 mL 8 M LiCl and four times with 10 mL H_2O, resuspended in 1 mL dH_2O and stored at -80 °C. The sacculi was later labelled with fluorescent fusion proteins to be visualised under a fluorescence microscope, as described in section 3.12.

3.11.1 Removal of teichoic acids

Removal of TAs from cell wall sacculi were prepared as described by Sanderson et al. (1962), however with some modifications (154). 100 μL of the isolated sacculi was centrifuged at 11,000 x g for 5 minutes. The sacculi was then resuspended in 1 mL of 10% (w/v) TCA. The tube was incubated at 37 ℃ with shaking for two weeks and immediately prepared for fluorescence microscopy.

3.12 Microscopy

The LSM 700 (Zeiss) laser scanning microscope with a 100x phase contrast objective was used to observe the binding patterns of the recombinant sfGFP-proteins against *S. dysgalactiae* cells and sacculi with and without TAs. The HPX 120 Illuminator (Zeiss) was used as a fluorescent light source, and the ORCA-Flash 4.0 digital complementary metal-oxide-semiconductor camera (Hamamatsu Photonics) attached to the microscope was used to capture images.

3.12.1 Labelling cells with sfGFP-proteins

The targets for protein attachment involved either a 2.5 mL culture of *S. dysgalactiae* strain MA201 at an OD_{550} = 0.25, 25 µL of isolated sacculi with teichoic acids or 1 mL of sacculi without TAs (treated with 10% TCA). The cells were then washed once in PBS and harvested at 4000 x g, while the sacculi (with and without TAs) was washed three times in PBS and collected at 11,000 x g. All targets were resuspended in 500 µL PBS-Tween (section 2.9.5) containing 15 µg/mL of the various sfGFP-$CbpD^{SD}$ versions. The mixture was then placed in the dark for 8 minutes at room temperature, immediately followed with three washing steps with PBS. Lastly, the pellets were dissolved in 150 µL PBS, and 0.5 µL of the *S. dysgalactiae* or sacculi (with or without TAs) cells marked with sfGFP-protein were applied within the wells of a microscope slide covered with a thin layer of 1.2% agarose (section 2.9.5). The microscope slide was cast with a thin layer of agarose in order to prevent movement of the bacteria cells and keep them immobilized. After applying the sample onto the wells, a cover glass was placed on top making the slide ready for fluorescence microscopy.

3.12.2 Fluorescence microscopy

The basic principle of fluorescence involves absorption of light at a specific wavelength, which causes the energy level of an electron to rise from a ground to an excited state. When the excited molecule returns to the ground state, it emits a photon of light of lower energy by a longer wavelength causing fluorescence. This therefore gives the fluorescent substance a distinct color when exposed to a specific wavelength, for instance UV light. While the phase-contrast microscopy uses phase shifts in light passing through the specimen, the fluorescence microscope uses a set of barrier filters where the illuminating light is passed through, and a dichroic mirror. The wavelength that excites the fluorescent dye passes only through the first filter, while the second filter blocks out this light and allows only the emitted fluorescence produced from the dye. The cells therefore glow against a dark background (113). The sfGFP reporter gene was used in this research for subcellular location of the target, and requires no substrate or co-factor to emit green light (112). It has an excitation maximum of 485 nm and an emission maximum of 510 nm, which gives it its inherent green fluorescence (155). Images were taken with an exposure time of ~1060 ms and at a wavelength of 485 nm.

3.12.3 Analysis of microscope images with ImageJ and MicrobeJ

Subcellular localization patterns were determined using ImageJ and its plugin software MicrobeJ (156, 157). MicrobeJ was used when analyzing the binding patterns of sfGFP-CbpD^{SD}, sfGFP-CbpD$^{\Delta SH3b}$ and the point-mutated versions. The bacteria were detected by several specifications: area (μM^2), length (μM), width, circularity, curvature, sinuosity, angularity, solidity and intensity, although the default settings [0-max] were mainly used as the bacteria could be detected using these values. In addition, exclude edges, shape descriptors, segmentation and chain were also options included in the description. After bacterial detection, the Maxima function was used to define the parameters for fluorescent foci detection, with the default setting yet again applied. Inside, outside, location, localization and coverage were the associations also included. An XYCellDensity plot from the Results Interface with a heatmap rendering was then used as a subcellular localization chart of the fluorescence foci relative to the center of the bacterial cell (158).

3.13 Growth curve assay

Optical density (OD) is the most commonly used method for quantifying the number of cells in a liquid suspension, and was used in this work to determine the growth of a bacterial culture at a specific wavelength. The principle is based on light scatter, or amount of intensity lost when light passes through an optical component, where wavelength 550 nm (OD_{550}) was used for *S. dysgalactiae* and wavelength 600 nm (OD_{600}) for *E. coli*. One disadvantage of optical density measurement is that it is only viable within a limited range (= 0.01 – 1.0) (159). The primary use of plate readers is OD measurement, yielding fast, inexpensive and readily automated measurements, where both a lysis-assay with CbpD^{G302A,G303A} and checking for immunity were performed. In addition to OD, the Sytox™ Green nucleic acid stain (Invitrogen™) was used as a viability assessment. This stain is a DNA-binding agent that cannot cross the cell membrane of living cells, and is therefore useful as an indication of lysis as the amount of fluorescence directly correlates with cell lysis by CbpD^{SD}. This dye becomes excited at 488 nm with an emission maximum of 523 nm, producing green fluorescence (160). The Hidex Sense (Hidex Oy) microplate reader was used in this project.

3.13.1 CbpD^{G302A,G303A} activity assay

CbpD^{SD} and CbpD^{G302A,G303A} were added to *S. dysgalactiae* grown in a 96-well polystyrene microtiter plate (Sarstedt), where lytic effect was measured by OD_{550} and Sytox™ fluorescence. The negative control of inactive enzyme was prepared by treating CbpD^{SD} with iodoacetamide, which alkylates and inactivates the catalytic cysteine residue found in the CHAP domain of CbpD^{SD} (section 1.3.1). 50 µg of CbpD^{SD} was firstly treated with 150 mM DTT and incubated for 5 minutes at room temperature. DTT is a reducing agent and was used to open the cysteine linkages in the protein. In order to prevent reformation of these disulfide linkages, IAM was subsequently used (161). Iodineacetamide to a final concentration of 0.5 M was added, and the mixture was incubated in the dark for 15 minutes.

3.13.2 Preparation for microplate reading

The appropriate *S. dysgalactiae* MA201 strain was grown in 5 mL C-medium (section 2.9.1) to OD_{550} = 0.5, and then diluted to OD_{550} = 0.05. Volumes of 280 µL of the diluted culture were transferred to twelve separate wells (280 µL x 12) in a 96-well microtiter plate containing 20 µL 5 mM Sytox™ Green. C- medium in a single well was used as reference. The plate was then incubated at 37 °C in the Hidex Sense plate reader, with the OD_{550} being measured automatically every 5 minutes. At OD_{550} = 0.2, CbpD^{SD} (a positive and a negative control after IAM treatment) and CbpD^{G302A,G303A} were added to each well to give final concentrations of 2.5 µg/mL, 1 µg/mL and 0.5 µg/mL. The microtiter plate was then placed back in the plate reader and continued for 200 cycles. Individual growth curves were made using the OD_{550} values and the Sytox™ Green values.

3.13.3 Immunity assay

S. dysgalactiae strains MM352 (015) and MM356 (047) (Table 2.3, section 2.1) were used to confirm the presence of an immunity mechanism against wildtype CbpD^{SD}. These strains were used, instead of the MA201 strain, as they have a functional ComRS-system. Preparations for the microplate reading was performed identical to what has been discussed above (section 3.13.2), although at OD_{550} = 0.2, ComS2 (section 2.6) was added to half of the wells of each strain in the 96-well microtiter plate to a final concentration 250 ng/mL. In twelve separate wells, where each strain occupied six wells each, half were induced with ComS2 and the rest left uninduced. After 20 minutes, CbpD^{SD} (final concentration of 1 µg/mL) was added to all wells. Similarly, individual growth curves were made using the OD_{550} values and the Sytox™ Green values.

4. Results

4.1 Overexpression and purification of sfGFP-CbpD$^{\Delta SH3b}$

CbpD has mostly been studied in *S. pneumoniae*, where its role in the fratricide mechanism is evident (72). Pneumococcal CbpD (CbpD^{SP}) binds specifically to the septal area of target cells where it makes cuts in the peptidoglycan resulting in cell lysis (64). It has also been shown that the SH3b domain of CbpD^{SP} is important for binding to peptidoglycan, but not for guiding the protein to the septum. This is accomplished by its choline binding domain which binds to choline moieties found on the teichoic acids in the cell wall of *S. pneumoniae* (73). The CbpD protein in *S. dysgalactiae* (CbpD^{SD}) contains a homologous SH3b domain, but instead of a choline binding domain, it has a conserved domain of unknown function (Figure 4.1). We hypothesize that this conserved domain in CbpD^{SD} is functionally analogous to the choline binding domain of CbpD^{SP}, i.e., it guides the protein to the division zone of target cells. In order to investigate this possible function of the conserved domain, several recombinant versions of CbpD^{SD} were overexpressed and purified (see section 3.9 for protocol).

Figure 4.1 Simplified illustrative overview of the CbpD^{SD} structure from *S. dysgalactiae*. The protein consists of a CHAP domain (in blue), an SH3b domain (in yellow) and the conserved domain (in grey). The CHAP and SH3b domains are similarly found in CbpD^{SP} but the protein has a conserved domain of unknown function instead of the choline binding domain of CbpD^{SP}. Created with BioRender (14).

It has been shown previously in the lab that purified sfGFP-CbpD^{SD}, in which the CHAP domain has been substituted with superfolder green fluorescent protein, binds specifically to the division zone of *S. dysgalactiae* target cells. To test if the conserved domain of CbpD^{SD} could have a similar function to the choline binding domain of CbpD^{SP}, namely that the domain alone localizes to the septum of target cells, we fused the sfGFP directly to the conserved domain for overexpression and purification. A 6xHis-tag was also fused to the proteins facilitating purification by using Ni^{2+} IMAC (section 3.9.3) (Figure 4.2).

Figure 4.2: The CHAP domain has been substituted with a superfolder green fluorescent protein (sfGFP) and fused directly to the conserved domain for overexpression and purification. This was done to test if the conserved domain alone guides CbpD^{SD} to the septum of target cells. A fused 6xHis-tag also facilitated purification. Created with BioRender (14).

The fusion-protein obtained from a 400 mL *E. coli* BL21 culture is shown in Figure 4.3A. A 280 nm absorbance peak was registered after 10 mL of linear gradient elution. Fractions 13-15 were therefore collected, which collectively had a protein yield of 2.92 mg/mL (NanoDrop™ 2000). Based on the SDS-PAGE and Coomassie blue staining, the purity of recombinant protein was in the range of ~75-80%. Additionally, and according to the SDS-PAGE, the protein construct had ~43 kDa sized-bands which is in accordance with its theoretical molecular mass of 43.57 kDa (Figure 4.3B).

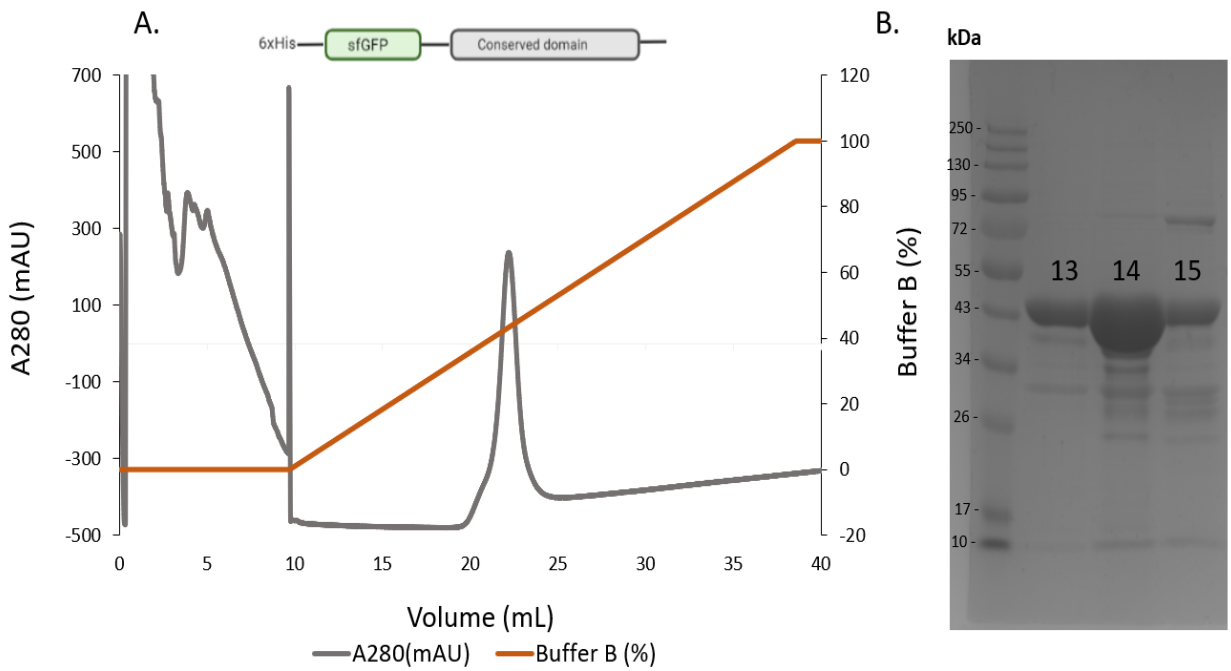

Figure 4.3: A. Chromatogram showing elution of sfGFP-CbpD$^{\Delta SH3b}$. The graph displays a 280 nm absorbance peak of the protein (grey line) after about 10 mL (~170 mM imidazole) of linear gradient elution with Buffer B (dark orange line). Fractions 13-15 were therefore collected. **B**. SDS-PAGE confirmed the presence of sfGFP-CbpD$^{\Delta SH3b}$ as correct band sizes (43.57 kDa) were displayed. The purity of the recombinant protein was estimated to be in the range of ~75-80%.

The sfGFP-CbpD^{SD} parental protein was similarly overexpressed and purified, in order to use the protein as control of normal localization of CbpD. Figure 4.4A displays the eluted fusion-proteins from a 400 mL culture, with 280 nm absorbance registered after 3-10 mL of the linear gradient elution of Buffer B. Fractions 8 and 9 were therefore collected. sfGFP-CbpD^{SD} was purified several times during the course of this work, and the protein yield therefore ranged from 0.15 mg/mL to 0.45 mg/mL. The SDS-PAGE and Coomassie blue staining of the recombinant protein indicated a protein purity of ~70%. Furthermore, the sfGFP-CbpD^{SD} has a theoretical molecular weight of 50.62 kDa, which was in accordance with the bands in the SDS-PAGE (Figure 4.4B).

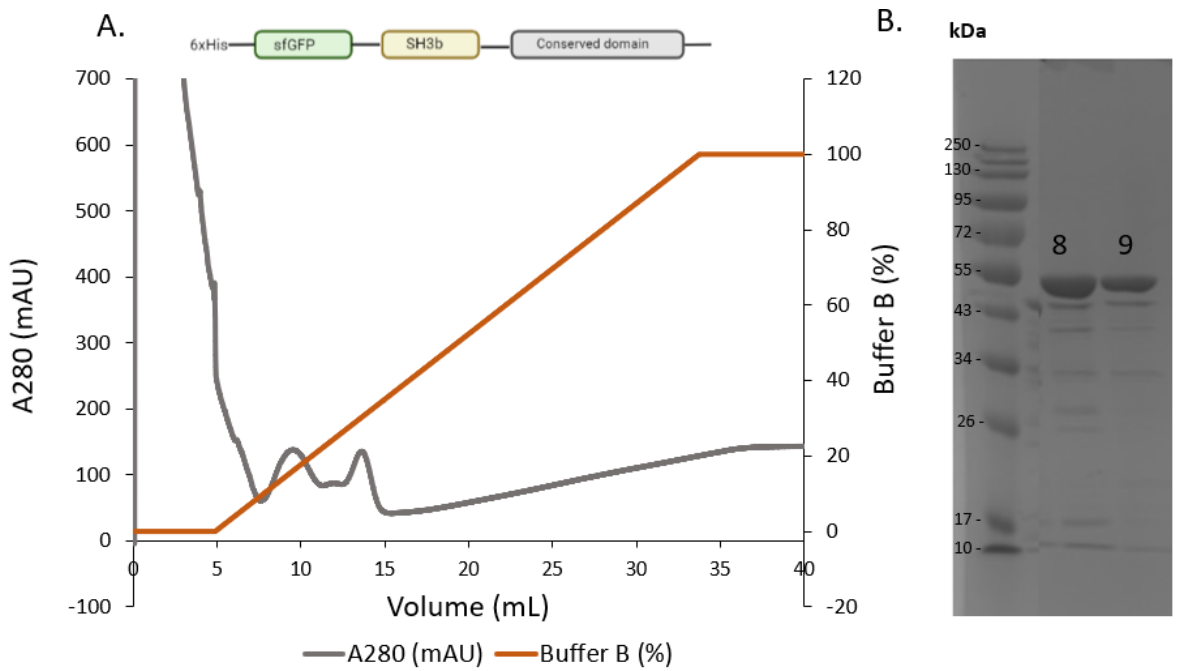

Figure 4.4: A. Chromatogram showing elution sfGFP-CbpD^{SD}, with the graph displaying 280 nm absorbance of the protein (grey line) after about 3 mL of linear gradient elution with Buffer B (~100 mM imidazole) (dark orange line). The first peak was assumed to be caused by contaminants, and only fractions 8 and 9 were collected. **B**. SDS-PAGE confirmed the presence of sfGFP-CbpD^{SD} as correct band sizes (50.62 kDa) were displayed, where the purity was estimated to be in the range of ~70%.

4.1.1 sfGFP-CbpD$^{\Delta SH3b}$ binds to the division zone

To test if the conserved domain alone could localize to the septum of dividing target cells, exponentially growing *S. dysgalactiae* strain MA201 cells were labelled with purified sfGFP-CbpD$^{\Delta SH3b}$ proteins (section 3.12) (Figure 4.5). For comparison with normal CbpD^{SD} localization, similar preparation was also performed for sfGFP-CbpD^{SD}. Fluorescence microscopy showed that the signal from sfGFP-CbpD$^{\Delta SH3b}$ was enriched at the division zone of cells, strongly resembling the localization pattern of sfGFP-CbpD^{SD}. All images were captured using the same settings (exposure time) on the microscope. XYCellDensity plots (created with ImageJ and its plug-in software MicrobeJ) confirmed that the conserved domain alone is sufficient to guide the protein to the septal area of target cells, as revealed by its strong density of subcellular foci at the outer edges of the bacterial cells (Figure 4.5, heatmap). The sfGFP-CbpD$^{\Delta SH3b}$ heatmap was made by analyzing 1375 bacterial cells (n=1375), whilst the sfGFP-CbpD^{SD} heatmap was made by analyzing 1272 bacterial cells (n=1272) using the same Maxima and XYCellDensity settings (section 3.12.3).

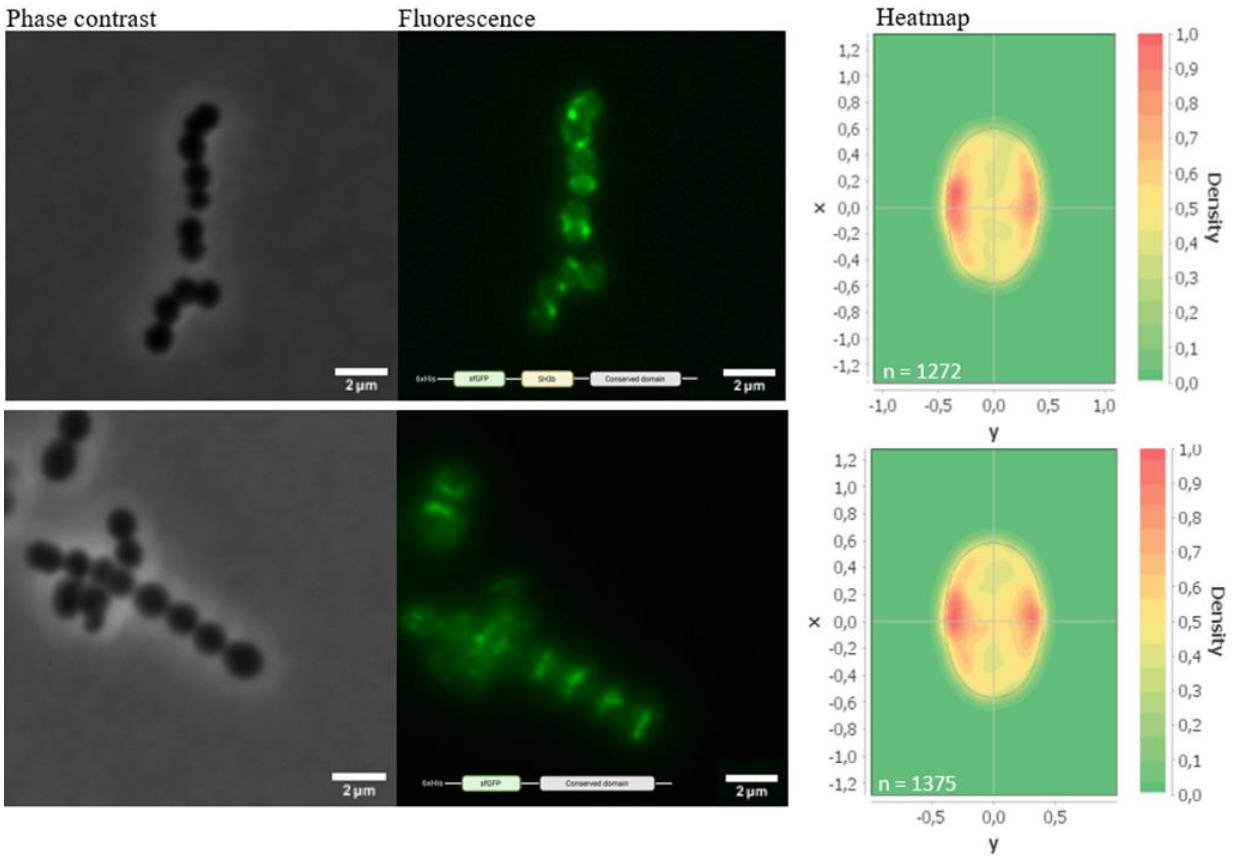

Figure 4.5: Subcellular localization of sfGFP-CbpD^{SD} and sfGFP-CbpD$^{\Delta SH3b}$ on target cells. The fluorescent images and heatmaps showed that both sfGFP-CbpD^{SD} and sfGFP-CbpD$^{\Delta SH3b}$ localized at the septal region of the *S. dysgalactiae* MA201 cells, indicating that the conserved domain alone is sufficient to guide the protein towards septum. The bar on the right of the heatmap specifies the level of subcellular location density, indicating that the two red areas of the heatmaps are high density levels of fluorescence foci (at the division zone). The axes of the cell density plot (x and y) also denotes the relative length. Scale bars for the microscopy images are 2 µm.

A linker region does exist between the sfGFP and SH3b domain (N-PSGEAS-C, Appendix 2) of the CbpD^{SD} proteins that potentially could be involved in division zone targeting. This region was not removed for any of the recombinant expressed proteins. Work by Eldholm et al. (2010) on CbpD^{SP} showed that the linker does not contribute in binding patterns (73). This revealed that the linker region does not possess a septal targeting ability, and conclusively it was believed similarly in this project. The conserved domain alone of CbpD^{SD} therefore guides the protein to the division zone of target cells.

4.1.2 Does sfGFP-CbpD$^{\Delta SH3b}$ bind to teichoic acids?

The CbpD^{SP} binds non-covalently to the phosphorylcholine moieties of wall- and lipoteichoic acids at the septal cell wall of *S. pneumoniae*, and it was therefore hypothesized that the CbpD^{SD} also binds to the teichoic acids found at the newly synthesized cell wall at the division zone of *S.*

dysgalactiae (64) (section 1.3.1). In order to confirm this, a protocol for successful removal of TAs on isolated sacculi had to be performed first. The removal of TAs is most efficiently done by incubating cell wall material in 48% hydrofluoric acid for 48 hours at 4 °C (73, 162). However, hydrofluoric acid is highly toxic, and for safety reasons, a different protocol was used in this work. There are other methods as well, for instance using 0.5 M NaOH, 1 M HCl or 10% TCA for removal of teichoic acids (154, 163, 164), where using NaOH and TCA on the isolated sacculi from *S. dysgalactiae* was performed in this project. Different conditions of treatment (i.e., temperature, light and duration) were also tested simultaneously. Subsequently, using 10% TCA on the sacculi for two weeks at 37 °C was redeemed as the better protocol for teichoic acid removal as other methods caused the sacculi to completely disappear. It has also been shown by Sanderson et al. that prolonged treatment of TCA (10%) removed 95% of *S. aureus* (strain Copenhagen) cell wall TAs (154). The purity of the peptidoglycan sacculi after two weeks of treatment remains unknown however, as residual TAs could remain after treatment.

The sfGFP-CbpD$^{\Delta SH3b}$ protein was labelled to the isolated sacculi from *S. dysgalactiae* cells without TAs before fluorescence microscopy analysis, where sacculi treated with water for two weeks was used as a control for sacculi containing TAs (Figure 4.6). It was seen that the sfGFP-CbpD$^{\Delta SH3b}$ bound to TA-enriched sacculi at its septal region, whereas sacculi deficient of teichoic acids were more evenly labelled with the GFP-fusion. This finding demonstrated that the conserved domain recognizes these peptidoglycan components and that they are needed for division zone recognition. Additionally, to determine whether the fluorescent signal for TCA-treated sacculi derived from sfGFP-CbpD$^{\Delta SH3b}$ or was background noise, unlabeled TCA-treated sacculi were imaged using the same microscopic parameters (Figure 4.6, bottom panel). The low fluorescent signal further confirmed that sfGFP-CbpD$^{\Delta SH3b}$ bound to sacculi without TAs. It does, however, not enrich at the septum. All images were captured using the same settings (exposure time) on the microscope and analyzed with the ImageJ software using the same brightness/contrast settings.

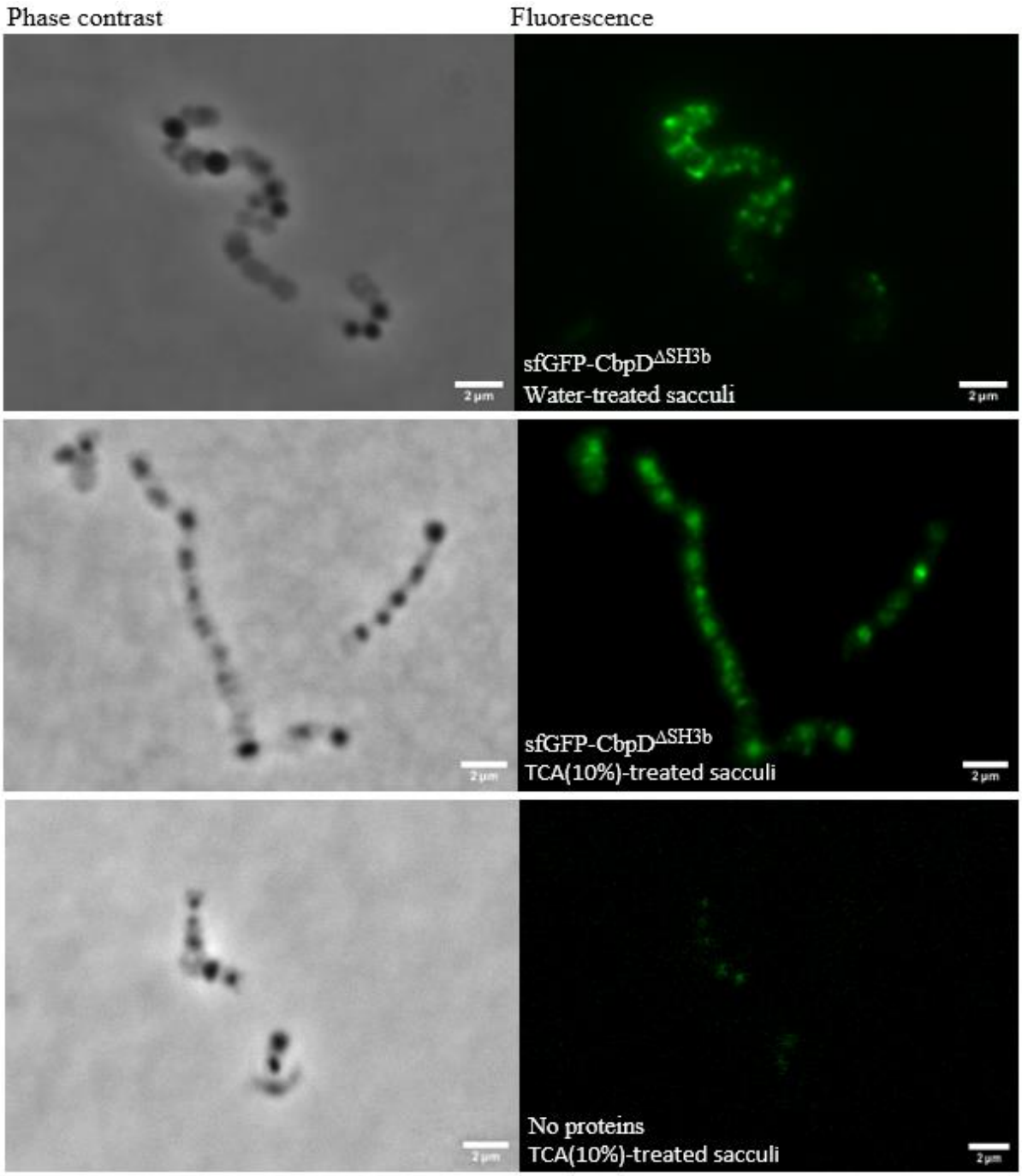

Figure 4.6: Phase contrast and fluorescence microscope images of sfGFP-CbpD$^{\Delta SH3b}$ labelled to two weeks of water-treated sacculi or two weeks of TCA(10%)-treated sacculi. The topmost panel is the control, followed with sfGFP-CbpD$^{\Delta SH3b}$ labelled to sacculi deficient of teichoic acids (removed by TCA). It was seen that sacculi deficient of TAs were more evenly labelled with the GFP-fusion. In order to exclude that the fluorescent signal from the fusion was background noise, unlabeled TCA-treated sacculi were imaged using the same microscopic parameters (bottom panel). Little fluorescent signal was registered from these images, and conclusively it was confirmed that sfGFP-CbpD$^{\Delta SH3b}$ bound to sacculi without TAs. It does however not enrich at the septum. Scale bars for the microscope images are 2 µm.

4.2 Identification of conserved amino acids by multiple sequence alignment

A multiple sequence alignment (MSA) was performed in order to identify conserved amino acids in the conserved domain of $CbpD^{SD}$. A selection of these conserved amino acids were individually mutated to amino acids of different charges within the sfGFP-$CbpD^{\Delta SH3b}$ sequence, subsequently creating different point-mutated sfGFP-$CbpD^{\Delta SH3b}$ protein versions. By visualizing them under the fluorescence microscope their binding patterns could be observed to reveal a conserved role of the conserved domain in directing the $CbpD^{SD}$ protein to the septal region of targeted cells. The $CbpD^{SD}$ amino acid sequence without its signal sequence (see Appendix 2) was aligned against CbpD from *S thermophilus*, *Streptococcus vestibularis*, *Streptococcus porcinus*, *Streptococcus uberis*, *Streptococcus iniae*, *Streptococcus ictaluri*, *Streptococcus equi* and *S. pyogenes*. These species were picked as their CbpD protein all harbor a homologous conserved domain (Figure 1.4, section 1.3.1), as well as possessing the ComRS regulatory system for regulation of competence (Figure 1.1, section 1.1). After entering the query sequence and specifying the search set in BLAST (140), the alignment with the highest query cover and lowest E-value were chosen. The aligned sequences of the various species were downloaded as FASTA files and uploaded to Clustal Omega in order to produce an MSA with an output format with character counts (165).

Figure 4.7 displays the MSA of the mentioned species CbpDs, but only of the conserved domain as this was of main significance (see Figure A.3 in Appendix 3 for the complete MSA). The protein alignment revealed that there are several conserved and highly similar amino acid residues across the different streptococcal species. The highly conserved leucine-alanine-glycine-glycine (LAGG) motif was identified, and since it is the only motif that was 100% conserved, it was of interest from the start of the project. Among the other highly conserved amino acids, numerous conserved glycine (G), as well as a couple of valines (V), prolines (P), phenylalanine (F) and leucines were observed in the aligned sequences. The glutamine (Q) and tryptophan (W) were interestingly only repeated once, and was therefore also chosen for further work in studying their importance for guiding the protein towards target cells' septum. These are, with the LAGG motif, highlighted in yellow in the following Figure 4.7.

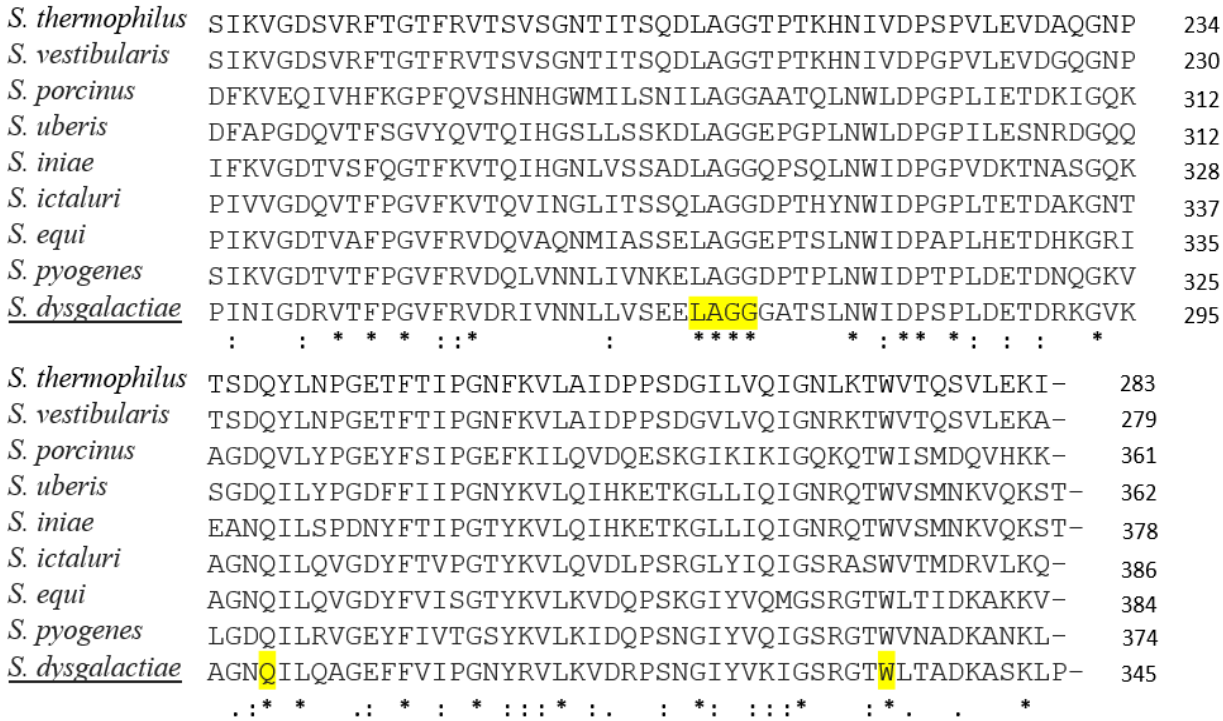

Figure 4.7: A multiple sequence alignment of the conserved domain found in CbpDs from various streptococcal species, including *S. dysgalactiae* as query. The following species were included in the MSA (with BLAST accession number also displayed): *S. thermophilus* (AKH33394.1**)**, *S. vestibularis* (WP_117648089), *S. porcinus* (WP_003083717), *S. uberis* (WP_203261572.1), *S. iniae* (WP_071127360), *S. ictaluri* (WP_008089638.1), *S. equi* (WP_012677198.1), *S. pyogenes* (WP_136094586.1). A conserved leucine-alanine-glycine-glycine (LAGG) motif and conserved amino acids glutamine (Q) and tryptophan (W) were chosen for further studies. The consensus symbols included mean: an asterisk (*) for a fully conserved residue, a colon (:) for highly similar residues and a period (.) for weak similar residues. Created with Clustal Omega (165).

The protein structure prediction services AlphaFold (166) and RoseTTAFold (167) were also used to make a prediction of the CbpD^{SD} protein by entering the amino acid sequence without its signal sequence. The 3D models were, to a certain degree, different, and the RoseTTAFold 3D prediction can therefore be found in Appendix 4 as Figure A.4. The AlphaFold prediction service was believed to be the most accurate, and is therefore presented in Figure 4.8. The predicated 3D structure of CbpD^{SD} showed a typical CHAP domain with an open cavity where the catalytical cysteine is located. The CHAP is connected to the SH3b domain via several loops and a single α-helix. Furthermore, the SH3b domain are predicted to be built up by six β-strands that form two β-sheets packed against each other forming the typical β-barrel fold found for these domains (75). The conserved domain is predicted to consist of eight anti-parallel β-strands that form a cylindric-like structure. However, this domain is probably predicted with less confidence, since no structural homologous exists in the databases. This is perhaps why some long stretches are predicted as

unstructured. Moreover, the conserved LAGG motif and amino acids Gln and Trp of the conserved domain are positioned, according to Figure 4.8, within the loops of these secondary protein structures and towards the surface of the protein. Being surface exposed suggests that the conserved motif and conserved amino acids have a role in directing the CbpD^{SD} protein towards the septal regions of the cellular division zone, which would conceivably be difficult if isolated within the domain core and away from the target.

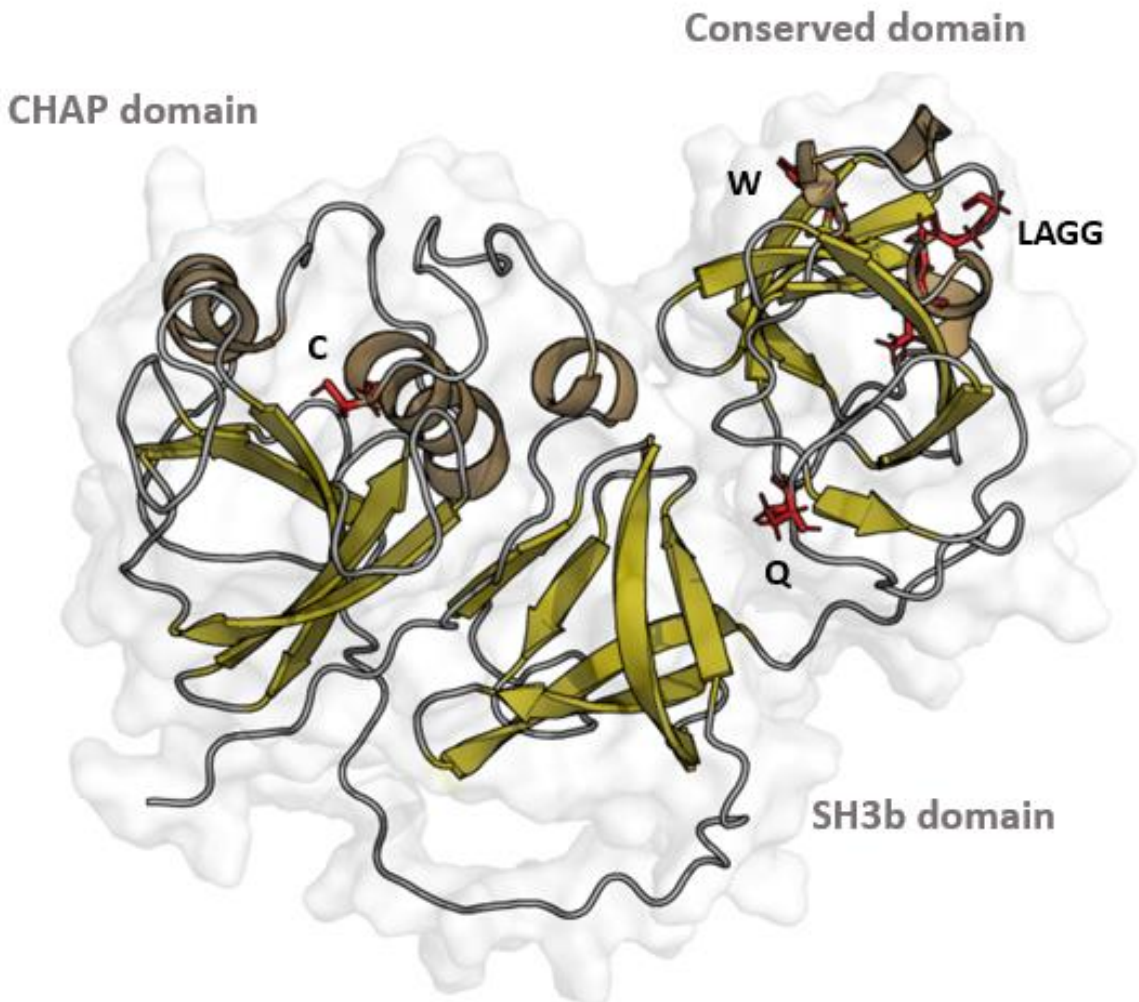

Figure 4.8: Predicted 3D structure of CbpD^{SD} of *S. dysgalactiae* strain MA201. The cysteine (C) of the CHAP domain, glutamine (Q), tryptophan (W) and the leucine-alanine-glycine-glycine motif (LAGG) of the conserved domain are the important conserved amino acids in this work and therefore presented as red sticks. Meanwhile, the predicted structure is presented as a cartoon with β-strands (yellow), α-helices (brown) and loops. A transparent surface behind was also included. Created with PyMOL (168).

4.3 Overexpression and purification of point-mutated sfGFP-CbpD$^{\Delta SH3b}$

The LAGG motif, Gln329 and Trp365 amino acids were chosen to mutate individually to produce different protein versions. This was made possible by constructing primers with 5' overhangs carrying site-directed mutations and then performing OE-PCR, as explained in sections 3.4.2 and 3.4.2.2. Both alanine and serine were introduced as the mutated amino acids, due to their non-bulky structure and charges. Ala is a hydrophobic amino acid, whilst Ser has a hydroxyl group making it polar but uncharged (150). Ala was therefore introduced as a mutation for the two last amino acids Gly (a neutral amino acid) of the LAGG motif, for Gln (has a polar uncharged side chain) and Trp (hydrophobic). Ser, on the other hand, was introduced as a mutation for the two first amino acids Leu (hydrophobic) and Ala (hydrophobic) of the LAGG motif. These were the preferred amino acids to use as they preferably do not impose structural limitations, but are physiochemically different from the original conserved amino acids. The exception is Trp which has the same charge as alanine, but is nonetheless less bulky. Four different point-mutated sfGFP-CbpD$^{\Delta SH3b}$ versions were successfully produced: a sfGFP-CbpD$^{\Delta SH3b,L300S,A301S}$, a sfGFP-CbpD$^{\Delta SH3b,G302A,G303A}$, a sfGFP-CbpD$^{\Delta SH3b,Q329A}$ and a sfGFP-CbpD$^{\Delta SH3b,W365A}$ (Figure 4.9). Appendix 5 includes the complete DNA sequences for each point-mutated sfGFP-CbpD$^{\Delta SH3b}$ version with the specific DNA changes needed to mutate the conserved amino acids into either Ala or Ser.

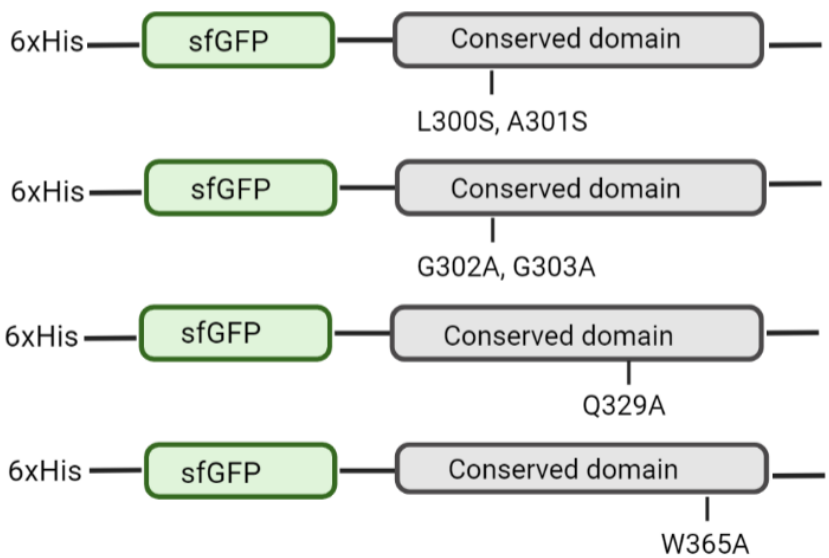

Figure 4.9: Schematic presentation of four sfGFP-CbpD$^{\Delta SH3b}$ versions with indicated point-mutations into either alanine or serine (L300S,A301S, G302A,G303A, Q329A and W365A). An N-terminal 6xHis tag was also fused to the proteins facilitating purification by Ni^{2+} IMAC.

The eluted point-mutated proteins from 400 mL cultures are shown in Figure 4.10 (A, C, E and G), with 280 nm absorbance registered roughly after 8-11 mL of linear gradient elution. Fractions were collected for each CbpD$^{\Delta SH3b}$-mutant, where sfGFP-CbpD$^{\Delta SH3b,\ L300S,\ A301S}$ had a protein yield of 0.40 mg/mL, sfGFP-CbpD$^{\Delta SH3b,G302A,G303A}$ a yield of 0.86 mg/mL, sfGFP-CbpD$^{\Delta SH3b,Q329A}$ with 1.50 mg/mL and sfGFP-CbpD$^{\Delta SH3b,W365A}$ with a protein yield of 0.55 mg/mL (NanoDrop™ 2000). Based on the SDS-PAGE and Coomassie blue staining, the purity of L300S,A301S mutant was in the range of ~70%, while for the other mutated proteins in the range of ~55-60%. The proteins had sizes in the ~43 kDa range (4.10B, D, F and H), which is in accordance with their theoretical molecular weights (also in a ~43 kDa range).

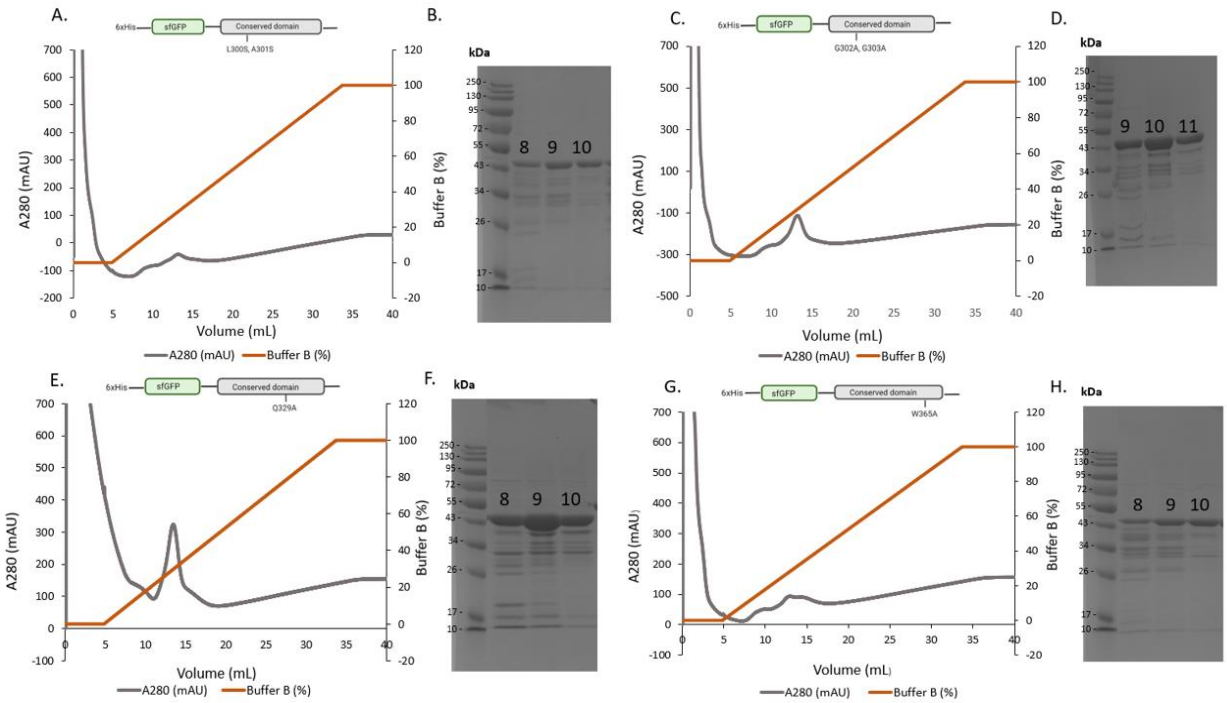

Figure 4.10: Chromatogram of sfGFP-CbpD$^{\Delta SH3b,L300S,A301S}$ (A), sfGFP-CbpD$^{\Delta Sh3b,G302A,G303A}$ (C), sfGFP-CbpD$^{\Delta SH3b,Q329A}$ (E) and sfGFP-CbpD$^{\Delta SH3b,W365A}$ (G). Proteins with a 280 nm absorbance peak (grey line) eluted at approximately 35% Buffer B (~175 mM imidazole, dark orange line). Fractions 8-11 were therefore collected. They were examined by SDS-PAGE (B, D, F and H) and confirmed to have a size corresponding to their theoretical mass, in the range of ~43 kDa. The purity of the recombinant proteins were also believed to be between 55-70%.

4.3.1 Identification of amino acids important for guiding CbpD^{SD} towards septum

Phase contrast and fluorescence microscope images were taken of *S. dysgalactiae* MA201 cells labelled with the different mutated sfGFP-CbpD$^{\Delta SH3b}$ versions to establish any differences in binding patterns. To be specific, would these mutated proteins still bind to the division zone of target cells? Binding to teichoic acids present at the septum was displayed in Figures 4.5 and 4.6,

and any difference in binding patterns presented here would suggest that these conserved amino acids have a conserved guiding role for $CbpD^{SD}$ towards the septal region. The sfGFP-$CbpD^{\Delta SH3b}$ protein was included as a positive control during microscopy. All images were captured using the same settings (exposure time) on the microscope and analyzed with ImageJ using the same brightness/contrast settings.

As seen in Figure 4.11, the sfGFP-$CbpD^{\Delta SH3b}$ control targets the division zone of MA201 target cells as expected. Collectively the binding is observably different for the point-mutated proteins, than the control. It was seen that septal recognition was lost for L300S,A301S, G302A,G303A, Q329A and W365A mutants, with an unevenly binding across the entirety of the targeted cells observed. This was similarly observed for their cognate heatmaps produced for each point-mutated sfGFP-fusion proteins, as the control clearly displayed strong density of subcellular location of the protein at septum (1493 bacteria analyzed). On the other hand, sfGFP-$CbpD^{\Delta SH3b,L300S,A301S}$ (n=1854), sfGFP-$CbpD^{\Delta SH3b,G302A,G303A}$ (n=1635), $CbpD^{\Delta SH3b,Q329A}$ (n=1805) and sfGFP-$CbpD^{\Delta SH3b,W365A}$ (n=1805) proteins showed strong density of subcellular fluorescent foci irregularly across the cells. Although, a higher degree of subcellular density at the septal regions seemed to be present for the two latter sfGFP-constructs (Q329A and W365A). Interestingly, small fluorescent foci were also observed within several of the cells across many of the fluorescent images. These were, though, only observable for the point-mutated protein versions and none in the control, with a general trend of occurring once or twice for each cell. This could perhaps be misfolded point-mutated sfGFP-$CbpD^{\Delta SH3b}$ aggregates bound strongly at certain points of the cellular wall.

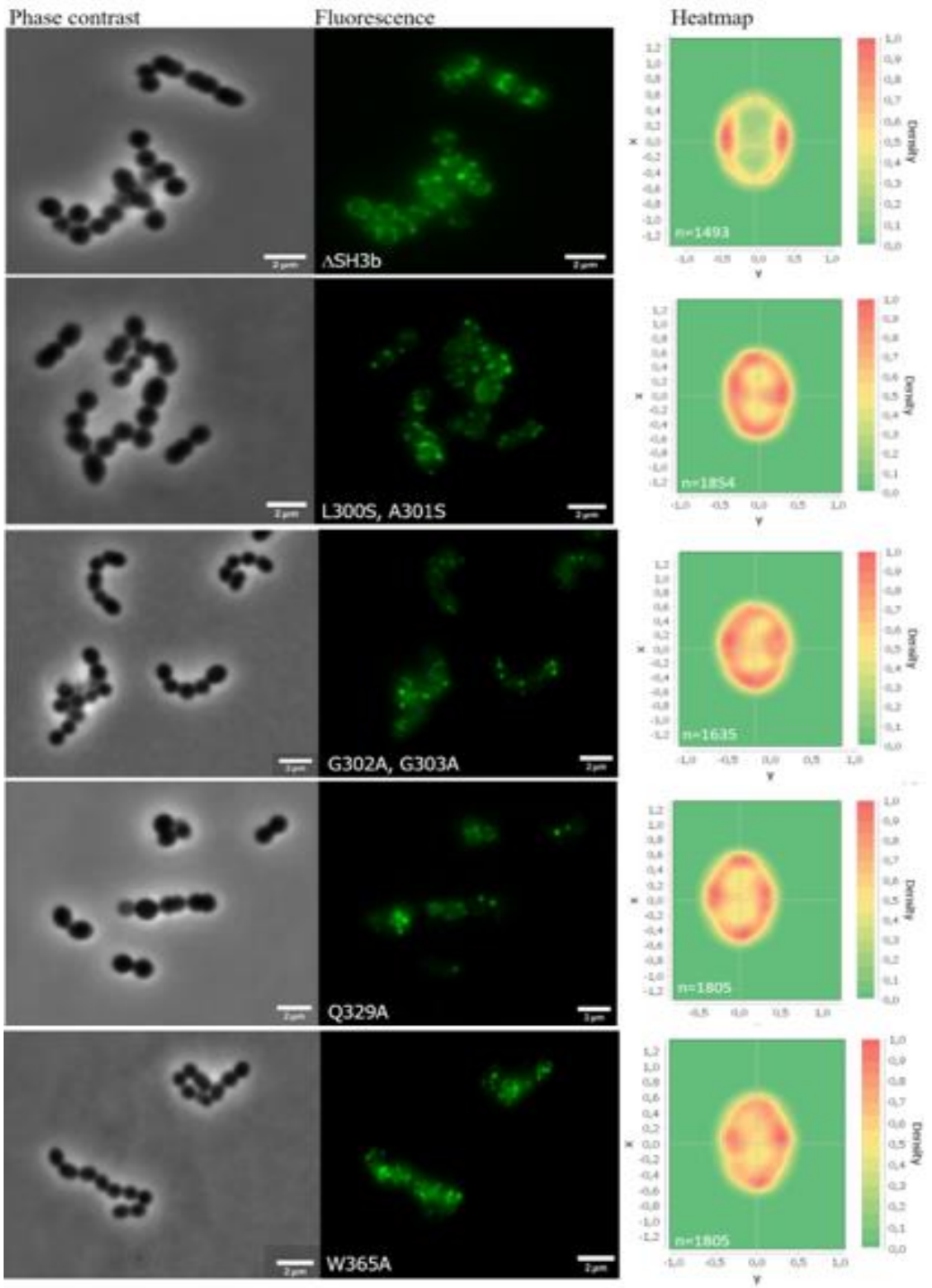

Figure 4.11: Phase contrast and fluorescent images of sfGFP-CbpD$^{\Delta SH3b}$ (control), sfGFP-CbpD$^{\Delta SH3b,L300S,A301S}$, sfGFP-CbpD$^{\Delta SH3b,G302A,G303A}$, sfGFP-CbpD$^{\Delta SH3b,Q329A}$ and lastly sfGFP-CbpD$^{\Delta SH3b,W365A}$. The microscopy images and heatmap of the control displayed septal binding, whereas the point-mutated sfGFP-versions displayed a lack of this with binding rather occurring unevenly across *S. dysgalactiae* MA201 target cells. This was also seen by their respective heatmaps, where n indicates the number of bacteria analyzed. Fluorescence foci were also observed in the mutated versions, where each speck seemed to occur once or twice per cells. Scale bars are 2 µM, with heatmaps produced by MicrobeJ.

4.4 Testing the lytic activity of CbpD^{G302A,G303A}

A growth curve assay was performed to measure the lytic activity of CbpD^{G302A,G303A} against *S. dysgalactiae* strain MA201 cells. Based on the results presented in Figure 4.11, sfGFP-CbpD$^{\Delta SH3b,G302A,G303A}$ was not enriched at septum suggesting that the LAGG motif is essential for the conserved domain's ability to guide the fratricin to the division zone. It was therefore hypothesized that CbpD^{G302A,G303A} would not bring about the CHAP domain to the cell septum to cause cell lysis. This point-mutated version of the protein was chosen to use in this assay, as it had high protein yield after purification (Figure 4.10 D) suggesting that the mutations had minimal effect on protein folding and due to significant interest in the highly conserved LAGG motif. CbpD^{G302A,G303A} was therefore a representative of a CbpD^{SD} version unable to bind correctly at the division zone of target cells, and the results from the assay was assumed to be similar if the other point-mutated sfGFP-fusions were used instead.

The full-length CbpD^{SD} protein with the CHAP domain, SH3b domain and the conserved domain was necessary (i.e., instead of sfGFP fused directly to SH3b) to measure its lytic effect on the *S. dysgalactiae* cells, and OE-PCR was subsequently used to introduce the G302A,G303A mutation to CbpD^{SD} (Figure 4.12). Wild-type CbpD^{SD} was used as positive control, as it is known to cause cellular lysis of targeted cells. As negative control we first wanted to use a version of CbpD^{SD} in which its catalytic cysteine (C65) residue in the CHAP domain was changed to Ala, rendering it inactive (Figure 4.12). The essentiality of the CHAP Cys residue for CbpD function was discovered by Håvarstein et al. (2006), who mutated the active site Cys into an Ala rendering it non-functional (70). Overexpression and purification of the C64A mutant was, however, unsuccessful (Appendix 6. Figure A.6.2). A wild-type CbpD^{SD} treated with iodoacetamide was used to create the negative control instead (see section 3.13.1).

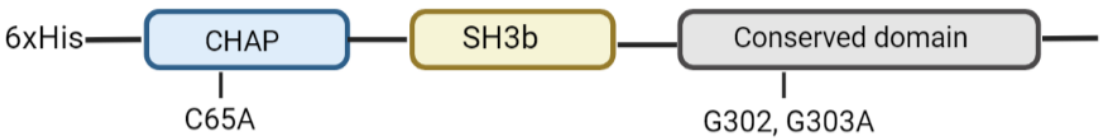

Figure 4.12: Schematic view of the full-length CbpD^{SD} where the positions of the point mutations C65A and G302A, G303A are indicated to create two individual proteins.

The CbpD^{G302A,G303A} protein (Figure 4.13) and positive control CbpD^{SD} were overexpressed and purified. Wild-type CbpD^{SD} was overexpressed in strain JA2, previously created in the research group, yielding 2.55 mg/mL of purified protein from a 400 mL cell culture (see Appendix 6, Figure A.6.1). Purified CbpD^{G302A,G303A} was obtained in the range of 0.42 mg/mL from the same culture volume (Figure 4.13A and B).

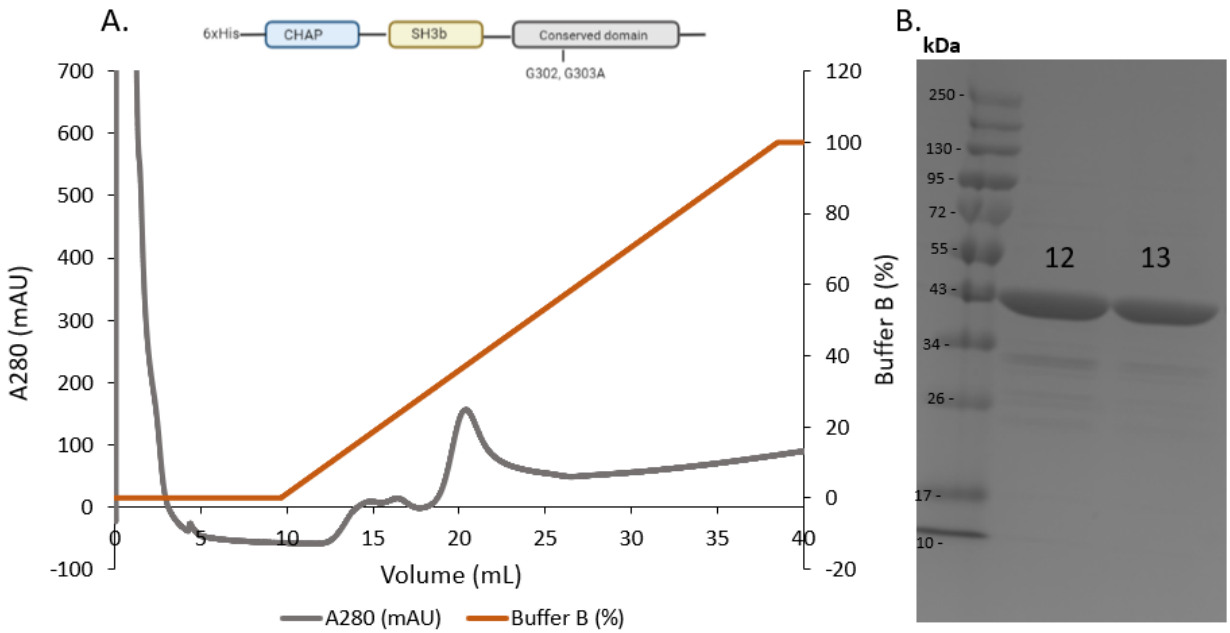

Figure 4.13: **A**. Chromatogram of CbpD^{G302A,G303A}. The graph displays absorbance at 280nm (grey line) after 12 mL of linear gradient elution and approximately 30% Buffer B (~150 mM imidazole, dark orange line). Fractions 12 and 13 were therefore collected. **B**. SDS-PAGE confirmed the presence of CbpD^{G302A,G303A} as correct band sizes (40.01 kDa) were displayed. The purity of the recombinant protein was believed to be in the range of ~85%.

Final concentrations of 0.5 µg/mL, 1 µg/mL and 2.5 µg/mL of CbpD^{SD}, CbpD^{G302A,G303A} and iodoacetamide-treated CbpD^{SD} were added to exponentially growing *S. dysgalactiae* cells at OD_{550} ≈ 0.2. Each concentration was carried out in three parallels. Sytox™ Green was also included for fluorescent measurements of cell lysis, as it is a more sensitive method to detect cell lysis compared to OD measurements, and could therefore detect potentially small lytic effects of the enzymes. It was seen that after the addition of the different CbpD^{SD} samples, the cells treated with wild-type enzyme lysed rapidly within 5 minutes as indicated by the heavy drop in OD_{550} and high fluorescent signal (Figure 4.14, only 1 µg/mL shown for simplicity of the figure). This showed that the wild-type CbpD^{SD} works efficiently against target cells. Addition of the mutated CbpD, having G302 and G303 changed to alanines, displayed significantly reduced ability to lyse target cells by only a small drop in OD_{550} and low fluorescent signal. Similarly, the iodoacetamide treated CbpD^{SD} had

no effect on the cells, which was expected. Although the LAAA mutation does not render the protein completely ineffective as some cell lysis was registered, the lytic effect was dramatically reduced suggesting that septal binding is critical for lytic activity. See appendix 7 (Figures A.7.1 and A.7.2) for the complete growth curve of all CbpD^{SD} samples at all concentrations.

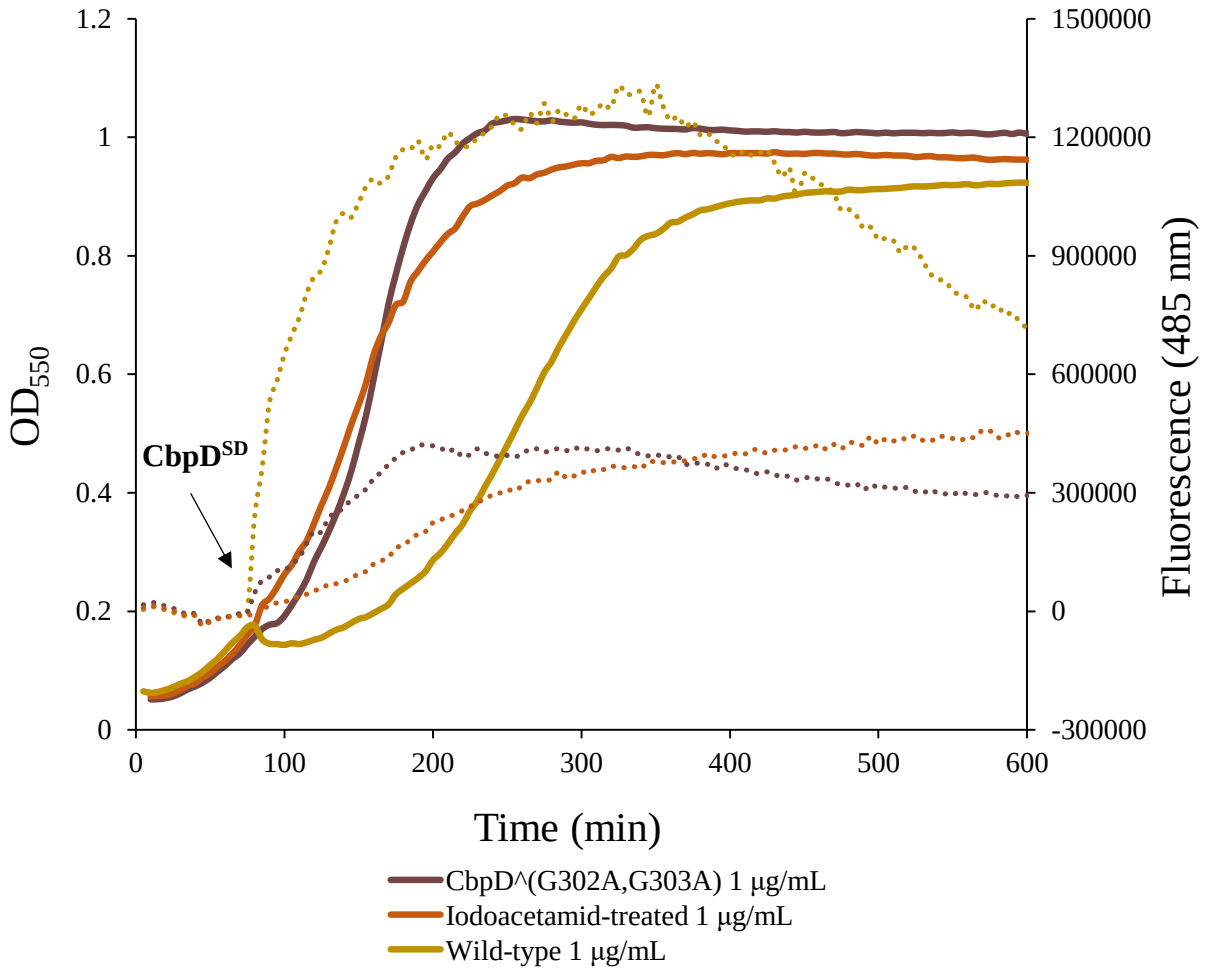

Figure 4.14: A lytic assay of *S. dysgalactiae* strain MA201 cells against CbpD^{SD} with G302A,G303A mutation. The cultures were incubated at 37 °C in C-medium, where growth was monitored until stationary phase by automatic detection of the OD_{550} at 5-minute intervals for 200 cycles. CbpD^{G302A,G303A} (1 µg/mL) was added at OD_{550} = 0.2 (shown by arrow). An iodoacetamide-treated CbpD^{SD} was used as negative control and a CbpD^{SD}-wild-type was used as positive control. It was seen that the mutant CbpD^{G302A,G303A} had nearly lost its ability to lyse the MA201 cells, as OD_{550} continued after adding the protein as well as registering low cell rupture of the cells by fluorescence (Sytox™ Green). Its conserved domain without a functioning LAGG motif can most likely not guide CbpD^{SD} towards septum, evidently showing that septal binding is necessary for hydrolytic activity of CHAP.

4.5 Competent *Streptococcus dysgalactiae* is immune against CbpD^{SD}

Håvarstein et al. (2006) set out to uncover the mechanism behind a clumping reaction produced from competent pneumococci, and discovered that a novel competence-induced protein called ComM was responsible. This protein was found to prevent the competent cell from committing suicide by its own fratricin, although the activity of ComM and which cellular processes it affects is still largely unknown (70, 71) (section 1.3.2). Such a fratricide immunity protein has not yet been described for other streptococci. But, having access to a pure and active fratricin from a streptococci in the pyogenic group, we wanted to use this enzyme to test if competent *S. dysgalactiae* also had a competence-induced mechanism giving them protection against their own fratricin. An obstacle for testing this has been that many *S. dysgalactiae* isolates vary in their abilities to enter competence in laboratory conditions. However, work done by PhD student Marita Torrissen Mårli in the Molecular Microbiology group have identified a *S. dysgalactiae* strain 047, which enters competence upon addition of ComS (XIP) inducer.

In order to check for an immunity, a growth curve assay was performed. *S. dysgalactiae* strain MA201 has a truncated ComR, and could therefore not be used in this assay as a functional ComRS-system for competence induction was necessary. *S. dysgalactiae* strain 015 modified in a way that *cbpD* is knocked out and replaced with a spectinomycin (MM352) resistance gene and *S. dysgalactiae* strain 047, as previously mentioned, were instead utilized (section 2.1, Table 2.3). Their ComR proteins are not truncated, and they both depend on ComS2 (QVDWWRL) for induction of the ComRS regulatory system. Each strain occupied six wells each of the 96-well polystyrene microtiter plates, half of which were induced with ComS2 (see section 3.13.3). After reaching an $OD_{550} \approx 0.2$, ComS2 (250 ng/mL) was added to the specified wells and after 20 minutes CbpD^{SD} wild-type (1 µg/mL) was added to all wells. A 20 min incubation with ComS2 was chosen to give the competent *S. dysgalactiae* cells enough time for induction of competence, expression of the potential immunity factor and to become fully immune before adding extracellular CbpD. Both OD_{550} and fluorescence (by Sytox™ Green) were measured every 5 minutes for 200 cycles, with the results from strain 015 displayed in Figure 4.15.

It was seen, after the addition of CbpD^{SD}-wild-type, that non-induced cells lysed strongly whereas the induced cells had no rupture. The non-induced cells had a strong drop in OD_{550} measurments, and high fluorescent signal indicating elevated amounts of cell rupture. The opposite was seen for

the cells induced to competence as only a small drop in OD_{550} was registered before continuing growing, and small amounts of cell rupture as displayed by the low fluorescent signal. There therefore seems to be an immunity mechanism produced in strain 015, which protects the competent cells from $CbpD^{SD}$ self-harm. Figure A.8 of Appendix 8 displays the results from strain 047. In contrast to strain 015, there is no clear difference between the induced competent streptococci and the non-induced by OD_{550} measurements nor fluorescence by Sytox™ Green.

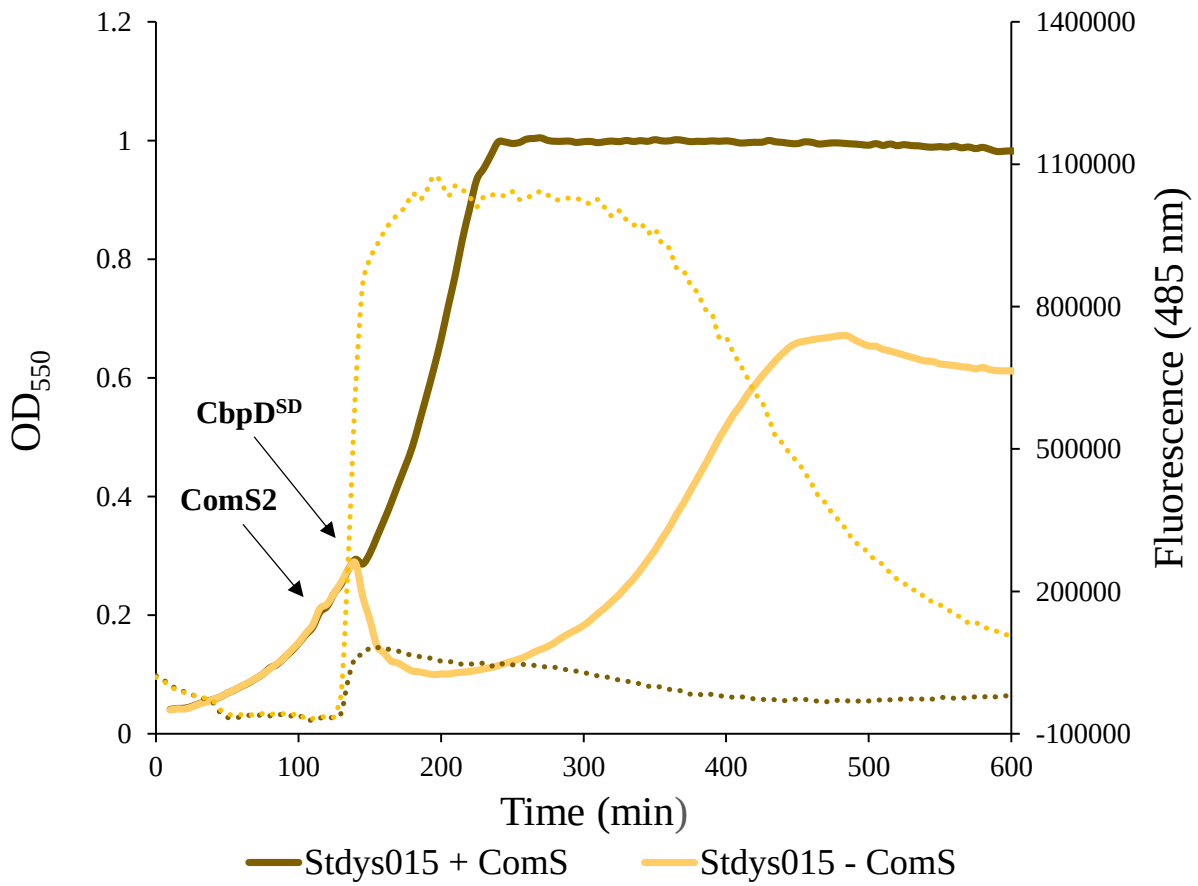

Figure 4.15: A growth curve of modified *S. dysgalactiae* strain 015 cells (*cbpD::spc*) where ComS2 (250 ng/mL) was added at OD_{550} = 0.2 to induce the cells to competence in half of the wells, and $CbpD^{SD}$-wild-type (1 µg/mL) added after 20 minutes. The cultures were incubated at 37 °C in C-medium, where growth was monitored until stationary phase by automatic detection of the OD_{550} at 5-minute intervals for 200 cycles. *S. dysgalactiae* strain 015 was proposed to have a immunity factor due to low OD_{550} drop and only small increase in fluorescence (Sytox™ Green) of the induced cells (in dark yellow). Non-induced cells were used as a control for lytic activity (in yellow). Similar results were obtained in repeated experiments.

5. Discussion

There are three mechanisms of horizontal gene transfer (transduction, conjugation and transformation), where genetic transformation has been explored in this project. The pathogenic *S. pneumoniae* is naturally competent for genetic transformation and has served as a model organism on this phenomenon. Its murein hydrolase $CbpD^{SP}$ is a crucial component of the pneumococcal fratricide mechanism, where its biological significance is immense as it readily makes DNA available for uptake and homologous integration into recipient genome. This provides the necessary material for further evolution and adaptability. Other streptococcal species outside the mitis group also contain gene sets that, in theory, should enable them to enter the competent state and execute fratricide. However, data showing active fratricide in streptococci outside the mitis group are limiting (51).

Members of the clinical relevant pyogenic subgroup of the *Streptococcus* genus have for a long time been considered non-transformable, though high recombination rates and HGT have been reported (41, 169). With the relatively recent discovery that *S. pyogenes* and other pyogenic species have a conserved set of competence genes and an *sigX* regulatory structure (for late competence gene expression), transformation could therefore have a larger impact on evolution and genome plasticity of the pyogenic group than previously believed (22). By identifying protein sequences with a CHAP domain and a SigX binding site in the promoter regions, several CbpD-like proteins from other streptococcal species (including *S. dysgalactiae*) have also been revealed. Their C-terminal halves are variable, although they most likely are functional analogues of $CbpD^{SP}$, where different recognition sites in their cellular envelope have been suggested (64). The $CbpD^{SD}$ has, however, a C-terminal conserved domain of unknown function. The present work has accordingly been an in-depth exploration of the conserved domain of $CbpD^{SD}$ in order to understand its biological function: does it localize to the septal cell wall, and does it bind to teichoic acids protruding from the peptidoglycan? The questions to be investigated could unravel the fratricide mechanism of *S. dysgalactiae*, its importance for making DNA available for integration and thereby genome plasticity for adaptation to changing environmental conditions.

5.1 The conserved domain guides $CbpD^{SD}$ to the cell division site

The binding pattern of recombinant sfGFP-$CbpD^{\Delta SH3b}$ protein on *S. dysgalactiae* MA201 cells revealed that the conserved domain from $CbpD^{SD}$ alone could recognize and bind to the septal cross

wall produced by the PBP2x/FtsW of the divisome machinery. This was clearly seen from the fluorescent images, as well as its cognate heatmap for subcellular location of the protein with high density at the septal locations (Figure 4.5, section 4.1.1). After these findings, identifying conserved amino acids which could constitute the septal guiding abilities of the protein was performed by creating a multiple sequence alignment of $CbpD^{SD}$ and various other streptococcal fratricins also with a conserved domain (Figure 4.7, section 4.2). A highly conserved LAGG motif and amino acids Gln329 and Trp365 were identified and individual proteins were made by site-directed mutations in order to observe protein localization. sfGFP-$CbpD^{\Delta SH3b,L300S,A301S}$, sfGFP-$CbpD^{\Delta SH3b,G302A,G303A}$, sfGFP-$CbpD^{\Delta SH3b,Q329A}$ and sfGFP-$CbpD^{\Delta SH3b,W365A}$ had significantly reduced abilities to bind to the division zone (Figure 4.11, section 4.3.1). Correct septal binding was also shown to be essential for the lytic function of $CbpD^{SD}$, since a $CbpD^{SD}$ version lacking the conserved LAGG motif ($CbpD^{G302A,G303A}$) displayed significant reduction in lytic abilities against sensitive *S. dysgalactiae* cells (Figure 4.14, section 4.4). Although this mutation (LAAA) does not render the protein fully ineffective, the low levels of cell rupture indicates that this motif might be essential for septum localization and thereby cause lysis of target cells by its CHAP domain. Since all point-mutated sfGFP-$CbpD^{\Delta SH3b}$ versions had a dramatically reduced ability for division zone binding, $CbpD^{G302A,G303A}$ was selected as a representative for them all. The conclusion is therefore drawn that the conserved LAGG motif and the Gln329 and Trp365 amino acids are critical for the localization of the $CbpD^{SD}$ protein to the septal regions, and that this localization is vital for the enzyme's ability to lyse the target cell. This is similar to what has been reported for the pneumococcal $CbpD^{SP}$.

The $CbpD^{SP}$ attacks the septal regions of target cells, but has a limited range as it can only target other species with choline-decorated teichoic acids at the division site (73). Similarly, it has been theorized that the conserved domain of $CbpD^{SD}$ most likely has a function equivalent to the CBD-domain of $CbpD^{SP}$, i.e., guiding the fratricin to the septal area of target cells. It has now been shown that $CbpD^{SD}$ targets the septal regions of *S. dysgalactiae* MA201 target cells and that teichoic acids are also necessary for the protein's ability for septal binding (Figure 4.6, section 4.1.2). The interesting question is why these specific murein hydrolases display such specificity towards the division zone. It was proposed by Eldholm et al. (2010) that the septal region is the weakest part of the cell wall, thereby allowing the CbpDs to cause cell rapture effortlessly. Straume et al. (2020) also discussed whether the peptidoglycan synthesized by the divisome is not perfect, and could

contain both gaps and small holes making this layer quite irregular and fragile. The CbpD-B6 (fratricin used in their work, from *S. mitis* strain B6) could use these imperfections from the otherwise tightly woven peptidoglycan to gain access to its substrate (73, 86). In addition, it was mentioned in section 1.4.1 that aPBPs synthesize a disordered secondary peptidoglycan layer, and although this two-layered septal cross wall has yet been demonstrated in *S. dysgalactiae*, it could make the peptidoglycan more robust and denser necessary for it to handle turgor pressure and retain the bacterial shape. The activity of aPBPs in the septal region also lags behind the primary divisome machinery, allowing CbpDs time to attack the weaker septal cross wall produced by the PBP2x/FtsW complex (86, 89). Although this model is based on results from studies in *S. pneumoniae*, due to the analogous function of $CbpD^{SD}$ and similar peptidoglycan synthesizing machineries, this might be similar for *S. dysgalactiae* as well. The septal cross wall might also be an indication of well-being, as cells that reproduce by cellular division displays that they thrive in the environmental conditions they are in. Although the biological role of these fratricins have not been determined, most data suggest that competent cells use them to acquire homologous DNA such that they can overcome challenges in the environment. By targeting healthy and closely-related cells carrying DNA capable of surviving stress-full environments, it makes sense to target their septal cross walls.

5.1.1 Identification of conserved amino acids important for septal binding of $CbpD^{SD}$

From the MSA (Figures 4.7 and A.3), several conserved amino acids were identified, but only the LAGG motif and amino acids Q329 and W365 were further studied. Creating individually point-mutated sfGFP-$CbpD^{\Delta SH3b}$ proteins of absolutely all these different conserved amino acids would have been ideal to map out their importance and significance at their specific locations. Nevertheless, only the above mentioned conserved amino acids were chosen due to time constraints. For instance, there are seven conserved glycine residues across the conserved domain. Figure 5.1 displays the position of the conserved glycine residues (in red and as sticks), where they are seen positioned within the loops towards the surface. Although this domain, as previously mentioned, probably is predicted with less confidence due to lack of homologous domains in the database, the several glycines can be hypothesized to have a structural role. Since Gly does not have a Cβ atom (i.e., no side chain) and can create many dihedral angle combinations, this allows for structural possibilities and flexibility of the protein. This could give the conserved domain of $CbpD^{SD}$ optimal function for binding to its substrate by adopting different confirmations (150,

170). Additionally, since the Gly is positioned between the β-strands within their turns due to their sterical flexibility, these promote the formation of antiparallel β-sheets. Two proline residues are also conserved and found in close proximity to each other, as seen on the MSA. In contrast to Gly, the Pro amino acids give structural rigidity to the protein. Perhaps a specific part of the conserved domain is rigid, while the rest of the domain has a higher degree of flexibility. It would therefore be interesting to make individual point-mutated sfGFP-CbpD^{SD} proteins where the several Gly's and Pro's are mutated into physiochemical different amino acids and see if the CbpD^{SD} loses its septal binding abilities and cause cellular lysis of target cells.

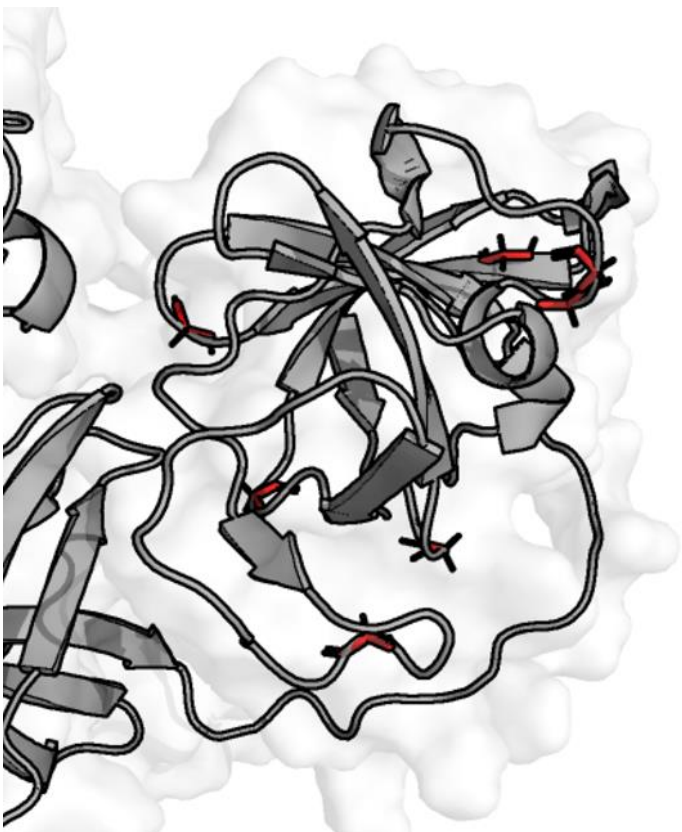

Figure 5.1: Predicted 3D structure of CbpD^{SD}, where the secondary structures are colored in grey (AlphaFold). The seven glycine residues are displayed as red sticks and can be seen within the loops. Created with PyMOL (168).

The function of the LAGG motif and Gln and Trp are unknown, but since they are positioned towards the surface (Figure 4.8, section 4.2) it can be speculated that they have a role in mediating binding to proteins or other molecules found at the septal regions, such as the teichoic acids. The motif is especially of interest as it consists of four amino acids conserved across several species with similar structure of CbpDs, and must proposedly have an essential functional role. It could be interesting to further research if the streptococcal species CbpD^{SD} is able to lyse has this LAGG-

motif, and if resistance species lack this. Previous work has revealed that CbpD^{SD} cannot fully lyse *S. pneumoniae* (105), where preliminary investigations has revealed that CbpD^{SP} does not contain this LAGG motif. Further work should be done, although this might suggest that LAGG recognizes something on the cells in common for all strain with LAGG-containing CbpDs. This would therefore constitute the binding specificity within the conserved domain of the protein, which can also be regarded for Q329 and W365, as their microscopy images were similar to that of the LAGG motif (Figure 4.11, section 4.3.1). Their heatmaps, however, indicate that the Q and W mutations still retain some septal binding abilities, with somewhat strong density of subcellular location in this area. Heatmaps produced from mutating the LAGG motif seem to have strong density of subcellular location across the entirety of the cells. This could indicate that the motif has the more important role of substrate recognition, with these amino acids (Gln and Trp) as supporting roles in cell wall binding specificity.

It should also be considered that introducing these mutations into the conserved domain could have caused the protein to misfold by affecting the stability of the protein. Proteins are dependent on their 3D structure, but alterations in this chain can cause the folding process to change, which gives rise to a misfolded protein (171). Its preferred fold or the native confirmation is lost, along with its ability, for instance, to bind to the septum and cause cellular lysis rupture for access to extracellular DNA. This could partly explain the several foci observed in Figure 4.11. They could be aggregates of misfolded proteins, where the lower signals could also be reasonable caused by the incorrect folding of GFP, as this would result in abnormally low fluorescence. However, since the mutated proteins were expressed and purified in fairly good amounts (0.4 – 1.5 mg/mL), it suggests that the proteins have an overall 3D structure that is correct or very close to native folding. Circular dichroism could have been applied to check that the secondary structures of the mutant proteins are similar to the native protein. Also, 3D crystallography or NMR studies would be necessary to confirm that the structure is unaltered. The amino acid residues chosen to mutate also are predicted (AlphaFold and RoseTTAFold) to lie towards the protein's surface and more so within the loops, and are probably more tolerant to changes within their sequences. Therefore, the belief is that the point-mutated sfGFP-CbpD$^{\Delta SH3b}$ versions have not misfolded and that something else is causing these observable foci. This has yet to be determined.

5.2 Teichoic acids in the septum function as a cell wall target recognition site for CbpD^{SD}

The binding patterns of sfGFP-CbpD$^{\Delta SH3b}$ to sacculi (with and without teichoic acids) differed significantly (Figure 4.6). The GFP-fusion was seen clearly bound to the septal regions of murein sacculi with TAs, whereas the protein labelled to sacculi deficient in TAs was evidently seen to have lost its septum binding ability. Although not enriched at the septum, the sfGFP-CbpD$^{\Delta SH3b}$ still bound to the sacculi. This latter finding was unexpected. The SH3b domain, as mentioned in section 1.3.1, most likely bind to the peptidoglycan part of the cell wall and forms a substrate-binding pocket together with the CHAP domain. Proteins with the SH3b domain are found in a variety of bacteria, where both the SH3b domain of CbpD^{SP} (73) and SH3b domain of ALE-1 (75) of *S. aureus* have been shown to bind to the peptidoglycan layer. Accordingly it can be speculated that the SH3b domain of CbpD^{SD} binds similarly. Eldholm et al. (2010) also suggested that it is not the interpeptide bridge that is recognized by the SH3b of CbpD^{SP}, but more likely either the Glc*N*AC or Mur*N*Ac residues (73). It was therefore interesting to see the Δ*SH3b* mutant remained bound to the *S. dysgalactiae* MA201 cellular envelope. Perhaps the conserved domain has a dual binding role, where it has the strongest affinity towards the TAs present at the division zone. When these TAs are lost, other parts of the cell wall of lower affinity could be the target for the conserved domain of CbpD^{SD}, although unspecific binding could also be a possibility. Another prospect is that a certain threshold of TAs at the septum is needed to recruit the conserved domain, and although the 10% TCA removed TAs, some of these polymers might still have remained causing sfGFP-CbpD$^{\Delta SH3b}$ to bind elsewhere on the sacculi. These are, however, only speculations. Interestingly, *S. thermophilus* (CbpDST) (Figure 1.4) lacks the SH3b domain (has only the N-terminal CHAP domain and a C-terminal conserved domain) but binds to the septal and polar regions of their targeted cells (Bjørnstad et al. (2012) (172). In this work, the MSA created included *S. thermophilus* (AKH33394.1) and showed that its conserved domain sequence is relatively homologous with the corresponding domain of CbpD^{SD}. It could therefore be hypothesized that CbpDST targets the division sone and the teichoic acids located there, although experimental evidence is lacking. Since it lacks the SH3b domain, the CbpDST might have evolved to bind to the peptidoglycan glycan chains or teichoic acids located outside the division zone, although with a stronger affinity for the TAs at the septum. It would therefore be interesting to conduct the same

experiment, as described in sections 3.11 and 3.11.1, on *S. thermophilus* and labelling the sacculi (with and without TAs) with sfGFP-CbpDST to see if unevenly binding of the cells is also seen.

The specific parts of the teichoic acids which the CbpD^{SD} targets are unknown. Research from previous lab work (105) showed that CbpD^{SD} (strain MA201) is able to lyse other streptococcal species from the mitis group (i.e., *S. gordonii* and *S. sanguinis*, furthest relatives of *S. pneumoniae*), species from the salivarius group (*S. thermophilus* and *S. salivarius*) and several species from the pyogenic group (*S. agalactiae* and *S. pyogenes* for instance). This target range would mean that these species possess specific recognition moieties on the TAs that CbpD^{SD} recognizes. It can be speculated that this moiety could be the D-alanine residue, as *S. dysgalactiae* (section 1.4.2), *S. gordonii* (173), *S. sanguinis* (174), *S. thermophilus* (175), *S. salivarius*, *S. agalactiae* and *S. pyogenes* (104) all have D-Ala decorated TAs. It would also be possible to quantify and compare the amount of D-alanine in the TAs of the different streptococci by liberating D-Ala from the teichoic acids and derivatizing them with a so-called Marfey's reagent (1,5-difluoro-2,4-dinitrobenzene). Reverse-phase chromatography easily separated D-Ala from L-Ala after derivatization (100, 176). Furthermore, it is known that the content of D-alanine esters is variable and depends on pH, salt concentrations (NaCl) and temperature (177). For instance, for *S. gordonii* LTAs, it has been reported that lower pH conditions cause an increase of *dlt* expression, while for *S. aureus* TAs (similar structure), an increase in pH, temperature or NaCl concentration cause a lower degree of D-alanylation where it has been shown that sublethal heating resulted in a 65% loss of the D-alanyl ester content (77, 178, 179). The *dlt* operon of *S. aureus* is also repressed by increasing the amount of divalent cations (such as Mg^{2+}) in the medium. It would therefore be interesting to alter the medium's salt concentration, pH or even the temperature during the growth of these mentioned streptococcal species and see if CbpD^{SD} still can cross-lysis. If CbpD^{SD} cannot, it may indicate that the D-alanine moieties at the TAs at the division site could be responsible. It was also suggested that the TAs of *S. pneumoniae* contain D-alanine residues, but by the works of a previous master student, CbpD^{SD} is not able to lyse *S. pneumoniae* (105). This might be due to the different teichoic acid structures these two species have, where *S. dysgalactiae* has a rhamnose polysaccharide (WTAs) and polyglycerol phosphate chains (LTAs) and *S. pneumoniae* uniquely of a tetrasaccharide-ribitol structure for both WTA and LTA. High structural variability of these polymers, even if they both contain D-Ala, might influence CbpD^{SD} not to cause lysis.

$CbpD^{SP}$ can target non-competent and closely-related species *S. oralis* and *S. mitis*, as they all have choline-decorated teichoic acids (51, 69). On the other hand, the phylogenetic relationship of $CbpD^{SD}$ targets is not that clear, as it was shown that the fratricin could cause cross-lysis of several strains belonging to almost all streptococcal groups (except the bovis and anginosus groups) (105). It was mentioned in section 1.3 that the biological function of the fratricide mechanism probably is to gain access to homologous DNA, as stable integration of diverging DNA sequences could be more harmful than beneficial due to functional incompatibility. The targets of fratricide should therefore naturally be more related species than species far unrelated, as they most likely possess more homologous DNA. HGT can, however, occur between distantly related organisms, but it is then more probable that genes belonging to simpler functional networks are transferred (than belonging to DNA replication, transcription or translation) (42). The fact that $CbpD^{SD}$ has an extensive target range could indicate that its recognition moiety on the TAs is quite general (such as the D-Ala) and not as limiting as choline. It is puzzling, however, that the $CbpD^{SD}$ can cause lysis of so many species and not only closely related ones, but perhaps its bovine habitat limits its range for cross-lysis of furthest relatives.

Why fratricins selectively bind at the septum is still an open question. It has been speculated that TAs inserted into the newly synthesized peptidoglycan layer undergoes chemical modifications, a kind of maturation, before they become part of the mature cell wall further away from the septum (99). *S. pneumoniae*, for example, express an enzyme called pce, which cleaves off a subfraction of choline residues on the teichoic acids inserted into the cell wall. It has been hypothesized that this could create a pattern of choline-rich teichoic acids near the septum. However, a Δ*pce* mutant is still sensitive to $CbpD^{SP}$, excluding pce's involvement in fratricin placement (unpublished data in the Molecular Microbiology group). Moreover, as already mentioned, one could also imagine that the concentration of teichoic acid is higher at the place of insertion (the septum), and that this locally high concentration of TAs has a titration effect recruiting fratricins to the division zone. Higher TA concentrations at the division zone have not been confirmed experimentally. What chemical determinants that induce the selectivity between peripheral and septal teichoic acids by these enzymes also remains to be determined. Curiously, the major autolysin in pneumococci called LytA also has a CBD, which it employs to bind teichoic acids before cutting the peptidoglycan. Similarly to $CbpD^{SP}$, this enzyme targets the TAs at the septum (99).

5.3 Substitution of cysteine to alanine in the CHAP domain of CbpD^{SD}

Creating the negative control for the CbpD^{SD} lysis assay (section 4.4) involved mutating the highly conserved cysteine residue within the CHAP domain to an alanine (C65A), which unfortunately did not succeed in this work. Attempts to overexpress and purify CbpD^{C65A} resulted in no protein yield, as seen in Figure A.6.2 in Appendix 6. The conserved Cys in the CbpD^{SD} is as expected predicted to be positioned in the active site cleft of the CHAP domain (Figure 4.8). The residue lies also more towards the core of the domain, and one could speculate that amino acid changes in this area is not so tolerated for the 3D structure of the domain. Changing this Cys to Ala has previously been done to inactivate CbpD^{SP} and PcsB (CHAP-containing protein involved in daughter cell separation in *S. pneumoniae*) (70). Changing the catalytic cysteine in the homologous CHAP domain of PcsB did not result in a misfolded protein (180). However, the cysteine mutated CbpD^{SP} was never tested for correct folding. It is therefore plausible to consider that the CbpD^{C65A} from *S. dysgalactiae* was misfolded upon overexpression. A small-scale experiment involving just overexpression of CbpD^{C65A} by IPTG and then directly onto the SDS-PAGE revealed no bands which could belong to the protein. The mutation (C to A) itself has most likely affected the expression of the protein. Exactly why the expression of CbpD^{C65A} failed in this work is not known.

5.4 Immunity against CbpD^{SD}

Clear experimental evidence of a fratricin immunity protein have only been found in pneumococci, which express the early competence protein ComM for protection against CbpD^{SP}. ComM homologs are found in other mitis streptococci, but not in more distantly related species such as *S. gordonii*, *S. thermophilus* and *S. dysgalactiae*. A key question is therefore whether the ComM-deficient streptococci express an alternative immunity factor. Experiments performed in this project have confirmed that *S. dysgalactiae* has a competence-induced immunity factor as competent *S. dysgalactiae* showed almost complete immunity against 1 µg/mL CbpD^{SD}. Non-competent cells displayed, on the other hand, high cellular lysis (Figure 4.15, section 4.5). To avoid lytic interference from self-produced CbpD^{SD} during competence induction, a Δ*cbpD* mutant (strain 015) was used in this experiment. *S. dysgalactiae* strain 047 was also included (Figure A.8, Appendix 8), but the resulting growth curve displayed similar OD_{550} and fluorescent measurments between the induced and non-induced cells. No clear indication that strain 047 produce immunity against CbpD^{SD} was therefore found. This could indicate differences between strains and isolates whether they are immune to CbpD activity or not. Although another possibility could be that the

047 cells have become induced for natural competence by various environmental factors before adding ComS2, rendering all cells immune. These are only speculations, and since it was observed that *S. dysgalactiae* strain 047 did not grow "smoothly" as shown by its OD_{550} measurments, more work should be done.

A MSA was also performed on the various CbpDs from *S. dysgalactiae* strains 015, 047 and MA201, with the following results presented in Figure A.9 in Appendix 9. It is seen that these CbpDs are highly conserved, with only a few divergences from the MA201 sequence. In the present work, it was therefore assumed that the $CbpD^{SD}$ from *S. dysgalactiae* strain MA201 would be equally efficient against *S. dysgalactiae* strains 015 and 047. Although strain 047 does contain a functional *cbpD* gene, and could potentially cause lysis by its native CbpD and provoke an immune response in avoidance of self-suicide, the CbpD level of expression within the isolate remained unknown. Knowing the concentration of the purified $CbpD^{SD}$ from MA201 and using a concentration that would undoubtedly cause lysis (1 µg/mL) was hence the best solution for this experiment. Optimally it would be preferred to test for immunity in strain MA201 but, as previously mentioned, this was not possible because it contained truncated ComR, leaving it non-competent upon ComS2 addition.

With the discovery that competent *S. dysgalactiae* strain 015 cells are immune to $CbpD^{SD}$, the competent cells must be protected from fratricide by an unknown mechanism. As discussed in section 5.2, *S. dysgalactiae* has a somewhat large target range where $CbpD^{SD}$ is able to lyse several species of several streptococci groups (105). The $CbpD^{SD}$ can thereby target non-competent cells, gain access to possibly homologous DNA and integrate into recipient genomes of the competent cells. Straume et al. (2017) discussed that ComM and $CbpD^{SP}$ probably interacts indirectly, where ComM could make changes to the cell wall structure of the newly synthesized peptidoglycan at the septal region. This could occur by modifications to the stem peptide in which the CHAP domain cleaves, the peptidoglycan to which SH3b binds or to the teichoic acids where the CBD domain binds to block the attachment of $CbpD^{SP}$. However, overexpression of ComM in *S. pneumoniae* produced cells with multiple septa to which a GFP-$CbpD^{SP}$ could still bind (71). Similarly, it has also been argued that ComM guarding and modifying only the division zone would be more beneficial than changing the entire cell, and since the ComM is only expressed a couple of minutes before $CbpD^{SP}$, it might not have time to protect the whole cell wall. These arguments, although

for *S. pneumoniae*, might be highly relevant for the immunity factor in *S. dysgalactiae* strain 015 as well. This could be a hypothesis for why $CbpD^{SD}$ targets the septum of non-competent cells, as the immunity factor guards or modifies only the septal area of competent cells. This is the first study showing that streptococci other than *S. pneumoniae* possess a competence-induced immunity mechanism against fratricins. Much therefore remains unknown and further research is needed. There has also been attempts to use the MEME Suite tool for motif-based sequence analysis (181) to search for other promoters with the same ComR early gene promoter binding site (AACAN GACA N_4 TGTCN TGTT N_{19} TATAAT, excluding the last part in the search) as the promoter regions of *comS* and *sigX* (section 1.2.2). A list of possible candidate motifs within the genome of *S. dysgalactiae* NCTC13762 was generated, with the P_{sigX} and P_{comS} on top. The following motifs further down the list contained an increasing number of mismatches to the consensus, and it was often seen that the suggested motif sites were in the middle of genes or belonged to genes whose activity were considered as not relevant for competence and transformation. The list of candidate motifs was quite long, and comparing one by one against the genome was deemed quite time-consuming and impossible to complete within the time frame of this work. Completion of this list provided by MEME Suite could possibly yield interesting findings of candidate genes with the consensus promoter region for ComR binding, assuming that this immunity gene is expressed early during competence. RNAseq comparing the transcriptomes of non-competent with competent *S. dysgalactiae* cells will also be done in the future to identify both early and late competence genes in this bacterium. Hopefully, one of the early genes can be identified as the immunity factor.

5.5 Is natural genetic transformation important for adaptability and evolution in the pyogenic group?

S. pyogenes and SDSE share similar ecological niches as they both infect humans with colonization in the respiratory tract, where similar disease associations have also been reported (35). It is therefore believed that entrance into a competence state and transformation for homologous recombination of extracellular DNA is relatively accessible for them due to overlapping habitats and a close evolutionary relationship. Although the HGT mechanism is unknown, it has been shown by Choi et al. (2012) high genetic flow between these two species by favoring *S. pyogenes* to SDSE replacing transfers (41). Additionally, Porcellato et al. (2021) reported that the majority of SDSE genetic content had high homology to the genes of *S. pyogenes* (30). For instance, the virulence gene *emm* encodes for the M-protein necessary to confer resistance to phagocytosis by

the host immune system, where *S. pyogenes* specific *emm*-genes have been found in SDSE isolates, revealing interspecies HGT events (32, 35). *S. pyogenes* and *S. dysgalactiae* does belong to the same ComRS pathway type II (section 1.2.3) and therefore have similar XIP sequences. There are however sequence variations among the XIP and ComR sequences, and competence induction between these two species might be XIP-specific. This would permit a subset of the bacterial population to become competent and allow for interspecies genetic exchange, just as there are different CSP pherotypes needed for a mixed population of predator and target cells among the mitis streptococci. However, it has been reported that both SDSE and *S. pyogenes* are low-specificity groups and hence respond to several XIP types. Research regarding XIP specificity still needs more work, and a possibility for SDSE/*S. pyogenes* responding to specie-specific XIPs by their ComR might still be the case (13). In addition, the use of penicillin (a β-lactam antibiotic) against *S. pneumoniae* infections over the last decades has resulted in penicillin-resistant pneumococci. The penicillin-binding proteins (section 1.4.1) are the targets for this antibiotic, but by expressing altered versions of these enzymes (low-affinity PBPs), higher penicillin concentrations are required to inhibit the enzymes. Furthermore, PBP2x, PBP2b and PBP1a are primary resistance determinants and commonly altered in resistant clinical isolates. Such alterations can come about by point-mutations of the transpeptidase domain or acquisition of low-affinity genes or gene fragments resulting in mosaic PBP genes, which have most likely occurred by transformation and homologous recombination events between closely related species (182–184). The PBPs of β-lactam-sensitive pneumococci are therefore relatively conserved, while low-affinity PBPs have high sequence diversity (182, 183). Fuursted et al. (2016) revealed four penicillin-resistant SDSE isolates with several amino acid substitutions within their transpeptidase domains of PBP2x, PBP1a and PBP1b. Two of the substitutions in PBP2x have commonly been observed in pneumococci and *S. agalactiae* penicillin-resistant strains (37). Low-affinity PBP mutations might have spread by natural transformation and recombination events to *S. dysgalactiae* from species of overlapping habitats, as this has been observed for *S. pneumoniae*. However, large segments of low-affinity PBPs genes, which seem more plausible to acquire from transformation events, have not yet been found. Acquisition of *pbp2b* and *pbp2x*- like genes, most likely from *S. dysgalactiae* donors to *S. pyogenes* by horizontal transfer, have also recently been reported (185).

SDSD is an animal-restricted pathogen and resides only in animals. Little genetic flow between SDSE to SDSD and SDSD to *S. pyogenes* is therefore found due to ecological mismatch (41). Porcellato et al. (2021) also reported that SDSD-specific genes have a closer resemblance to homologs in other animal-associated pathogens, for instance to *S. agalactiae*. A *lac2* operon, a part of a mobile genetic element, constitutes a selective growth advantage in lactose-rich environments (e.g., milk) for *S. agalactiae*, which was ubiquitous in bovine SDSD-isolates perhaps indicating important environmental adaptation (30). These two species have overlapping ecological habitats, providing ideal opportunities for genetic recombination for insertion of non-mobile and mobile genetic elements. Although Porcellato et al. concluded that cross-species transmissions are rare and that these subspecies are equipped with distinct repertories contributing to specific host adaptation, zoonosis has been observed by appearance of zoonotic diseases in humans (30, 186, 187). This could perhaps allow genetic exchange by competence and transformation (and fratricide) between SDSD and human streptococcal species, further developing its host adaptivity in humans and virulence of a human SDSD pathogen.

In section 1.1.1, it was mentioned that it is not known if *S. dysgalactiae* employs natural competence and genetic transformation to acquire genes for adaptability and evolution. It has been suggested that competence is regulated in a strict manner in the pyogenic group and population mixing might reflect exchange via transformation (22). From the findings in this thesis, *S. dysgalactiae* possesses a functional $CbpD^{SD}$, which is believed to help the competent cells getting access to homologous DNA for uptake and recombination. This murein hydrolase binds to the teichoic acids located at the septal region of non-competent cells, where the competent cells (for *S. dysgalactiae* strain 015) are protected by a supposed immunity mechanism. Competence for genetic transformation therefore seems like a plausible mechanism of HGT, especially because many *S. dysgalactiae* strains contain all genes required to become competent and to execute fratricide. With the given examples above for SDSE and SDSD strains and isolates who have gone through genetic exchange for evolution (for instance, in virulence or avoidance of antibiotics) and adaptability (into new hosts), transformation should not be discounted as a method of genetic exchange between pyogenic species, and may be more or equally important as transduction and conjugation.

6. Conclusion and future research

S. dysgalactiae subspecies *dysgalactiae* is an important animal pathogen and is primarily associated with clinical and subclinical bovine mastitis in cows, sheep and goats, which strongly plagues the dairy industry. SDSD can also cross-infect from animals into humans, most often in individuals in the fishing or animal industry, but can nonetheless potentially be a serious risk to immunocompromised patients. Antibiotics must be used to treat these infections, and in the prospect of sharing the same fate as penicillin-resistant *S. pneumoniae*, resistant isolates might be a possibility as well (29, 38, 186, 187). Penicillin-resistant SDSE isolates have now recently been recorded (37). The acquisition of virulence factors, resistance genes and adaptability to new hosts is dependent on the different horizontal gene transfer mechanisms, where transformation is often discounted. Several discoveries have been made throughout the present work, suggesting that natural transformation and fratricide could play a more significant role in HGT among *S. dysgalactiae* than previously assumed. Studying the $CbpD^{SD}$ from an SDSD strain has revealed many similarities with $CbpD^{SP}$ from *S. pneumoniae,* a key protein for gene exchange in this species. These similarities include possessing a C-terminal domain (CBD/conserved domain) which directs the fratricin to the septal regions, binding to the TAs present at the division zone, efficient lysis on exponentially growing cells and that competent producer cells express a fratricin immunity factor. A LAGG motif in the conserved domain of $CbpD^{SD}$ was shown to be essential for its septal binding ability, and hence necessary for target cell lysis by the CHAP domain. It is therefore a strong belief that natural competence for genetic transformation is a crucial mode of HGT for this species of streptococci, and perhaps for many other streptococci where transformation has previously been discounted. In this manner, genome analyses have recently suggested that mobile genetic elements are shared between *S. dysgalactiae* strains for host adaptability (30). It will be interesting to see in the future if there is a link between the bacteria's ability to share these elements and their ability to become natural competent. This work has increased our knowledge of an otherwise barely characterized fratricin in *S. dysgalactiae*, leaving a foundation for further research on fratricide, natural competence and genetic transformation as well as the immunity mechanism in this species.

There are several prospects for further work, for instance, testing the binding and effect of $CbpD^{SD}$ on teichoic acid knock-down mutants by utilizing the CRISPRi system (a technique for repression

of gene expression) and purifying CbpD^{SD} for crystallography in order to determine the protein 3D structure. Due to time constraints, these two research areas were not possible to pursue. This could have made it possible to study the levels of TAs necessary for CbpD^{SD} to bind at septum and see the location of the conserved amino acids on a very precise 3D structure of the fratricin made by X-ray crystallography. This 3D structure would also reveal how the CHAP domain and the SH3b domain are positioned against each other and therefore how they could work together, as the 3D structure from AlphaFold/RoseTTAFold was predicted with various confidence across the structure. A co-crystallization of extracted WTAs with CbpD^{SD} could also possibly reveal how this protein binds to the TAs. Other exciting prospects could be to identify the specific cut site for the CHAP domain, along with the specific TA recognition molecule the conserved domain of CbpD^{SD} is guided towards along the division zone. Furthermore, understanding the immunity mechanism discovered in *S. dysgalactiae* strain 015 would also be interesting to further research. The transcriptomic approach RNAseq should be utilized to find competence-induced genes (by comparing transcriptomes of non-competent with competent *S. dysgalactiae* cells), where an immunity gene would hopefully be among them. Additionally, only two strains (015 and 047) were used in this experiment, and including other strains should be considered if the immunity experiment were to be repeated. Finally, finding a natural competent strain for genetic manipulations (e.g., knock-out experiments) would be ideal, and although this investigation is ongoing in the Molecular Microbiology lab, it should be a strong focus for the coming future.

Conclusively, there are several opportunities for further research on the CbpD^{SD} fratricin, which would provide essential knowledge of CbpD from the pyogenic group, and where this knowledge might be of importance for the future in understanding the spread of antibiotic resistance genes and zoonotic capabilities this species potentially possess.

References

1. Nobbs AH, Lamont RJ, Jenkinson HF. 2009. *Streptococcus* Adherence and Colonization. Microbiol Mol Biol Rev 73:407–450.

2. Hardie J m., Whiley R a. 1997. Classification and overview of the genera *Streptococcus* and Enterococcus. Journal of Applied Microbiology 83:1S-11S.

3. Patterson MJ. 1996. *Streptococcus*, p. . *In* Baron, S (ed.), Medical Microbiology, 4th ed. University of Texas Medical Branch at Galveston, Galveston (TX).

4. Granum PE. 2015. Matforgiftning - smitte gjennom mat og vann. 4. edition. Cappelen Damm.

5. Abranches J, Zeng L, Kajfasz J, Palmer S, Chakraborty B, Wen Z, Richards V, Brady L, Lemos J. 2018. Biology of Oral Streptococci. Microbiol Spectr 6.

6. Willenborg J, Goethe R. 2016. Metabolic traits of pathogenic streptococci. FEBS Letters 590:3905–3919.

7. Postma PW, Lengeler JW, Jacobson GR. 1993. Phosphoenolpyruvate:carbohydrate phosphotransferase systems of bacteria. Microbiol Rev 57:543–594.

8. Lancefield RC. 1933. A serological differentiation of human and other groups of hemolytic streptococci. J Exp Med 57:571–595.

9. Facklam R. 2002. What Happened to the Streptococci: Overview of Taxonomic and Nomenclature Changes. Clin Microbiol Rev 15:613–630.

10. Richards VP, Palmer SR, Pavinski Bitar PD, Qin X, Weinstock GM, Highlander SK, Town CD, Burne RA, Stanhope MJ. 2014. Phylogenomics and the Dynamic Genome Evolution of the Genus *Streptococcus*. Genome Biol Evol 6:741–753.

11. Sherman JM. 1937. The streptococci. Bacteriol Rev 1:3–97.

12. Okura M, Osaki M, Nomoto R, Arai S, Osawa R, Sekizaki T, Takamatsu D. 2016. Current Taxonomical Situation of *Streptococcus suis*. Pathogens 5:45.

13. Shanker E, Morrison DA, Talagas A, Nessler S, Federle MJ, Prehna G. 2016. Pheromone Recognition and Selectivity by ComR Proteins among *Streptococcus* Species. PLoS Pathog 12:e1005979.

14. BioRender. BioRender. https://biorender.com/. Retrieved 3 February 2022.

15. Krzyściak W, Pluskwa KK, Jurczak A, Kościelniak D. 2013. The pathogenicity of the *Streptococcus* genus. Eur J Clin Microbiol Infect Dis 32:1361–1376.

16. Bhowmick R, Maung N, Hurley BP, Ghanem EB, Gronert K, McCormick BA, Leong JM. 2013. Systemic disease during *Streptococcus pneumoniae* acute lung infection requires 12-lipoxygenase-dependent inflammation. J Immunol 191:10.4049/jimmunol.1300522.

17. Zadoks RN, Middleton JR, McDougall S, Katholm J, Schukken YH. 2011. Molecular Epidemiology of Mastitis Pathogens of Dairy Cattle and Comparative Relevance to Humans. J Mammary Gland Biol Neoplasia 16:357–372.

18. von Wintersdorff CJH, Penders J, van Niekerk JM, Mills ND, Majumder S, van Alphen LB, Savelkoul PHM, Wolffs PFG. 2016. Dissemination of Antimicrobial Resistance in Microbial Ecosystems through Horizontal Gene Transfer. Frontiers in Microbiology 7.

19. Vekemans J, Gouvea-Reis F, Kim JH, Excler J-L, Smeesters PR, O'Brien KL, Van Beneden CA, Steer AC, Carapetis JR, Kaslow DC. 2019. The Path to Group A *Streptococcus* Vaccines: World Health Organization Research and Development Technology Roadmap and Preferred Product Characteristics. Clinical Infectious Diseases 69:877–883.

20. Engholm DH, Kilian M, Goodsell DS, Andersen ES, Kjærgaard RS. 2017. A visual review of the human pathogen *Streptococcus pneumoniae*. FEMS Microbiology Reviews 41:854–879.

21. Salvadori G, Junges R, Morrison DA, Petersen FC. 2019. Competence in *Streptococcus pneumoniae* and Close Commensal Relatives: Mechanisms and Implications. Front Cell Infect Microbiol 0.

22. Mashburn-Warren L, Morrison DA, Federle MJ. 2010. A novel double-tryptophan peptide pheromone is conserved in mutans and pyogenic Streptococci and Controls Competence in *Streptococcus mutans* via an Rgg regulator. Mol Microbiol 78:589–606.

23. Vandamme P, Pot B, Falsen E, Kersters K, Devriese Lay 1996. Taxonomic Study of Lancefield Streptococcal Groups C, G, and L (*Streptococcus dysgalactiae*) and Proposal of *S. dysgalactiae* subsp. *equisimilis* subsp. nov. International Journal of Systematic and Evolutionary Microbiology 46:774–781.

24. Vieira VV, Teixeira LM, Zahner V, Momen H, Facklam RR, Steigerwalt AG, Brenner DJ, Castro AC. 1998. Genetic relationships among the different phenotypes of *Streptococcus dysgalactiae* strains. Int J Syst Bacteriol 48 Pt 4:1231–1243.

25. Farrow JAE, Collins MD. 1984. Taxonomic Studies on Streptococci of Serological Groups C, G and L and Possibly Related Taxa. Systematic and Applied Microbiology 5:483–493.

26. Alves-Barroco C, Caço J, Roma-Rodrigues C, Fernandes AR, Bexiga R, Oliveira M, Chambel L, Tenreiro R, Mato R, Santos-Sanches I. 2021. New Insights on *Streptococcus dysgalactiae* subsp. *dysgalactiae* Isolates. Frontiers in Microbiology 12.

27. Jensen A, Kilian M. 2012. Delineation of *Streptococcus dysgalactiae*, Its Subspecies, and Its Clinical and Phylogenetic Relationship to *Streptococcus pyogenes*. J Clin Microbiol 50:113–126.

28. Rato MG, Bexiga R, Florindo C, Cavaco LM, Vilela CL, Santos-Sanches I. 2013. Antimicrobial resistance and molecular epidemiology of streptococci from bovine mastitis. Veterinary Microbiology 161:286–294.

29. Cheng WN, Han SG. 2020. Bovine mastitis: risk factors, therapeutic strategies, and alternative treatments — A review. Asian-Australas J Anim Sci 33:1699–1713.

30. Porcellato D, Smistad M, Skeie SB, Jørgensen HJ, Austbø L, Oppegaard O. 2021. Whole genome sequencing reveals possible host species adaptation of *Streptococcus dysgalactiae*. Sci Rep 11:17350.

31. Yeruham I, Schwimmer A, Braverman Y. 2000. Epidemiological, bacteriological and economical aspects of mastitis associated with yellow jacket wasps (Vespula germanica) in dairy cattle herd.

32. Hughes JM, Wilson ME, Brandt CM, Spellerberg B. 2009. Human Infections Due to *Streptococcus dysgalactiae* Subspecies *equisimilis*. Clinical Infectious Diseases 49:766–772.

33. Wajima T, Morozumi M, Hanada S, Sunaoshi K, Chiba N, Iwata S, Ubukata K. 2016. Molecular Characterization of Invasive *Streptococcus dysgalactiae* subsp. *equisimilis*, Japan. Emerg Infect Dis 22:247–254.

34. Oppegaard O, Mylvaganam H, Kittang BR. 2015. Beta-haemolytic group A, C and G streptococcal infections in Western Norway: a 15-year retrospective survey. Clinical Microbiology and Infection 21:171–178.

35. McNeilly CL, McMillan DJ. 2014. Horizontal gene transfer and recombination in *Streptococcus dysgalactiae* subsp. *equisimilis*. Front Microbiol 5:676.

36. Biedenbach DJ, Toleman MA, Walsh TR, Jones RN. 2006. Characterization of fluoroquinolone-resistant beta-hemolytic *Streptococcus* spp. isolated in North America and Europe including the first report of fluoroquinolone-resistant *Streptococcus dysgalactiae* subspecies *equisimilis*: report from the SENTRY Antimicrobial Surveillance Program (1997-2004). Diagn Microbiol Infect Dis 55:119–127.

37. Fuursted K, Stegger M, Hoffmann S, Lambertsen L, Andersen PS, Deleuran M, Thomsen MK. 2016. Description and characterization of a penicillin-resistant *Streptococcus dysgalactiae* subsp. *equisimilis* clone isolated from blood in three epidemiologically linked patients. Journal of Antimicrobial Chemotherapy 71:3376–3380.

38. Chennapragada SS, Ramphul K, Barnett BJ, Mejias SG, Lohana P. A Rare Case of *Streptococcus dysgalactiae* Subsp. *Dysgalactiae* Human Zoonotic Infection. Cureus 10:e2901.

39. Soucy SM, Huang J, Gogarten JP. 2015. Horizontal gene transfer: building the web of life. 8. Nat Rev Genet 16:472–482.

40. Avery OT, MacLeod CM, McCarty M. 1944. Studies on the chemical nature of the substance inducing transformation of pneumococcal types. J Exp Med 79:137–158.

41. Choi SC, Rasmussen MD, Hubisz MJ, Gronau I, Stanhope MJ, Siepel A. 2012. Replacing and Additive Horizontal Gene Transfer in *Streptococcus*. Molecular Biology and Evolution 29:3309–3320.

42. Thomas CM, Nielsen KM. 2005. Mechanisms of, and Barriers to, Horizontal Gene Transfer between Bacteria. Nat Rev Microbiol 3:711–721.

43. Griffith F. 1928. The Significance of Pneumococcal Types. Epidemiology & Infection 27:113–159.

44. Shen C-H. 2019. Nucleic Acids: DNA and RNA, p. 1–25. *In* Diagnostic Molecular Biology. Elsevier.

45. Sia RHP, Dawson MH. 1931. In vitro transformation of pneumococcal types. J Exp Med 54:701–710.

46. Johnston C, Martin B, Fichant G, Polard P, Claverys J-P. 2014. Bacterial transformation: distribution, shared mechanisms and divergent control. 3. Nature Reviews Microbiology 12:181–196.

47. Håvarstein LS, Coomaraswamy G, Morrison DA. 1995. An unmodified heptadecapeptide pheromone induces competence for genetic transformation in *Streptococcus pneumoniae*. Proc Natl Acad Sci U S A 92:11140–11144.

48. Claverys J-P, Prudhomme M, Martin B. 2006. Induction of competence regulons as a general response to stress in Gram-positive bacteria. Annu Rev Microbiol 60:451–475.

49. Johnsborg O, Håvarstein LS. 2009. Regulation of natural genetic transformation and acquisition of transforming DNA in *Streptococcus pneumoniae*. FEMS Microbiology Reviews 33:627–642.

50. Stevens KE, Chang D, Zwack EE, Sebert ME. Competence in *Streptococcus pneumoniae* Is Regulated by the Rate of Ribosomal Decoding Errors. mBio 2:e00071-11.

51. Straume D, Stamsås GA, Håvarstein LS. 2015. Natural transformation and genome evolution in *Streptococcus pneumoniae*. Infection, Genetics and Evolution 33:371–380.

52. Prudhomme M, Attaiech L, Sanchez G, Martin B, Claverys J-P. 2006. Antibiotic Stress Induces Genetic Transformability in the Human Pathogen *Streptococcus pneumoniae*. Science 313:89–92.

53. Moreno-Gámez S, Sorg RA, Domenech A, Kjos M, Weissing FJ, van Doorn GS, Veening J-W. 2017. Quorum sensing integrates environmental cues, cell density and cell history to control bacterial competence. 1. Nat Commun 8:854.

54. Havarstein LS, Diep DB, Nes IF. 1995. A family of bacteriocin ABC transporters carry out proteolytic processing of their substrates concomitant with export. Molecular Microbiology 16:229–240.

55. Håvarstein LS, Gaustad P, Nes IF, Morrison DA. 1996. Identification of the streptococcal competence-pheromone receptor. Molecular Microbiology 21:863–869.

56. Ween O, Gaustad P, Håvarstein LS. 1999. Identification of DNA binding sites for ComE, a key regulator of natural competence in *Streptococcus pneumoniae*. Molecular Microbiology 33:817–827.

57. Morrison DA, Lee MS. 2000. Regulation of competence for genetic transformation in *Streptococcus pneumoniae*: a link between quorum sensing and DNA processing genes. Research in Microbiology 151:445–451.

58. Håvarstein LS, Hakenbeck R, Gaustad P. 1997. Natural competence in the genus *Streptococcus*: evidence that streptococci can change pherotype by interspecies recombinational exchanges. J Bacteriol 179:6589–6594.

59. Carrolo M, Pinto FR, Melo-Cristino J, Ramirez M. 2009. Pherotypes are driving genetic differentiation within *Streptococcus pneumoniae.* BMC Microbiology 9:191.

60. Johnsborg O, Kristiansen PE, Blomqvist T, Håvarstein LS. 2006. A Hydrophobic Patch in the Competence-Stimulating Peptide, a Pneumococcal Competence Pheromone, Is Essential for Specificity and Biological Activity. J Bacteriol 188:1744–1749.

61. Fontaine L, Boutry C, de Frahan MH, Delplace B, Fremaux C, Horvath P, Boyaval P, Hols P. 2010. A Novel Pheromone Quorum-Sensing System Controls the Development of Natural Competence in *Streptococcus thermophilus* and *Streptococcus salivarius*. J Bacteriol 192:1444–1454.

62. Gardan R, Besset C, Guillot A, Gitton C, Monnet V. 2009. The Oligopeptide Transport System Is Essential for the Development of Natural Competence in *Streptococcus thermophilus* Strain LMD-9. Journal of Bacteriology 191:4647–4655.

63. Fontaine L, Goffin P, Dubout H, Delplace B, Baulard A, Lecat-Guillet N, Chambellon E, Gardan R, Hols P. 2013. Mechanism of competence activation by the ComRS signalling system in streptococci. Molecular Microbiology 87:1113–1132.

64. Berg KH, Biørnstad TJ, Johnsborg O, Håvarstein LS. 2012. Properties and Biological Role of Streptococcal Fratricins. Appl Environ Microbiol 78:3515–3522.

65. Laurenceau R, Péhau-Arnaudet G, Baconnais S, Gault J, Malosse C, Dujeancourt A, Campo N, Chamot-Rooke J, Le Cam E, Claverys J-P, Fronzes R. 2013. A Type IV Pilus Mediates DNA Binding during Natural Transformation in *Streptococcus pneumoniae*. PLoS Pathog 9:e1003473.

66. Claverys J-P, Martin B, Polard P. 2009. The genetic transformation machinery: composition, localization, and mechanism. FEMS Microbiol Rev 33:643–656.

67. Steinmoen H, Knutsen E, Håvarstein LS. 2002. Induction of natural competence in *Streptococcus pneumoniae* triggers lysis and DNA release from a subfraction of the cell population. Proc Natl Acad Sci U S A 99:7681–7686.

68. Berg KH, Ohnstad HS, Håvarstein LS. 2012. LytF, a Novel Competence-Regulated Murein Hydrolase in the Genus *Streptococcus*. J Bacteriol 194:627–635.

69. Claverys J-P, Martin B, Håvarstein LS. 2007. Competence-induced fratricide in streptococci. Mol Microbiol 64:1423–1433.

70. Håvarstein LS, Martin B, Johnsborg O, Granadel C, Claverys J-P. 2006. New insights into the pneumococcal fratricide: relationship to clumping and identification of a novel immunity factor. Mol Microbiol 59:1297–1307.

71. Straume D, Stamsås GA, Salehian Z, Håvarstein LS 2017. Overexpression of the fratricide immunity protein ComM leads to growth inhibition and morphological abnormalities in *Streptococcus pneumoniae*. Microbiology 163:9–21.

72. Kausmally L, Johnsborg O, Lunde M, Knutsen E, Håvarstein LS. 2005. Choline-Binding Protein D (CbpD) in *Streptococcus pneumoniae* Is Essential for Competence-Induced Cell Lysis. J Bacteriol 187:4338–4345.

73. Eldholm V, Johnsborg O, Straume D, Ohnstad HS, Berg KH, Hermoso JA, Håvarstein LS. 2010. Pneumococcal CbpD is a murein hydrolase that requires a dual cell envelope binding specificity to kill target cells during fratricide. Mol Microbiol 76:905–917.

74. Eldholm V, Johnsborg O, Haugen K, Ohnstad HS, Håvarstein LS 2009. Fratricide in *Streptococcus pneumoniae*: contributions and role of the cell wall hydrolases CbpD, LytA and LytC. Microbiology 155:2223–2234.

75. Lu JZ, Fujiwara T, Komatsuzawa H, Sugai M, Sakon J. 2006. Cell Wall-targeting Domain of Glycylglycine Endopeptidase Distinguishes among Peptidoglycan Cross-bridges. Journal of Biological Chemistry 281:549–558.

76. Silhavy TJ, Kahne D, Walker S. 2010. The Bacterial Cell Envelope. Cold Spring Harb Perspect Biol 2:a000414.

77. Neuhaus FC, Baddiley J. 2003. A Continuum of Anionic Charge: Structures and Functions of d-Alanyl-Teichoic Acids in Gram-Positive Bacteria. Microbiol Mol Biol Rev 67:686–723.

78. Vollmer W, Blanot D, De Pedro MA. 2008. Peptidoglycan structure and architecture. FEMS Microbiology Reviews 32:149–167.

79. Rajagopal M, Walker S. 2017. Envelope Structures of Gram-Positive Bacteria. Curr Top Microbiol Immunol 404:1–44.

80. Kohanski MA, Dwyer DJ, Collins JJ. 2010. How antibiotics kill bacteria: from targets to networks. Nat Rev Microbiol 8:423–435.

81. Vollmer W, Massidda O, Tomasz A. 2019. The Cell Wall of *Streptococcus pneumoniae*. Microbiol Spectr 7.

82. Severin A, Tomasz A. 1996. Naturally occurring peptidoglycan variants of *Streptococcus pneumoniae*. J Bacteriol 178:168–174.

83. Barreteau H, Kovac A, Boniface A, Sova M, Gobec S, Blanot D. 2008. Cytoplasmic steps of peptidoglycan biosynthesis. FEMS Microbiol Rev 32:168–207.

84. Pinho MG, Kjos M, Veening J-W. 2013. How to get (a)round: mechanisms controlling growth and division of coccoid bacteria. 9. Nat Rev Microbiol 11:601–614.

85. Sauvage E, Kerff F, Terrak M, Ayala JA, Charlier P. 2008. The penicillin-binding proteins: structure and role in peptidoglycan biosynthesis. FEMS Microbiology Reviews 32:234–258.

86. Straume D, Piechowiak KW, Olsen S, Stamsås GA, Berg KH, Kjos M, Heggenhougen MV, Alcorlo M, Hermoso JA, Håvarstein LS. 2020. Class A PBPs have a distinct and unique role in the construction of the pneumococcal cell wall. Proc Natl Acad Sci U S A 117:6129–6138.

87. Mohammadi T, van Dam V, Sijbrandi R, Vernet T, Zapun A, Bouhss A, Diepeveen-de Bruin M, Nguyen-Distèche M, de Kruijff B, Breukink E. 2011. Identification of FtsW as a transporter of lipid-linked cell wall precursors across the membrane. EMBO J 30:1425–1432.

88. Berg KH, Stamsås GA, Straume D, Håvarstein LS. 2013. Effects of Low PBP2b Levels on Cell Morphology and Peptidoglycan Composition in *Streptococcus pneumoniae* R6. Journal of Bacteriology 195:4342–4354.

89. Straume D, Piechowiak KW, Kjos M, Håvarstein LS. 2021. Class A PBPs: It is time to rethink traditional paradigms. Molecular Microbiology 116:41–52.

90. Szwedziak P, Löwe J. 2013. Do the divisome and elongasome share a common evolutionary past? Curr Opin Microbiol 16:745–751.

91. Zapun A, Vernet T, Pinho MG. 2008. The different shapes of cocci. FEMS Microbiol Rev 32:345–360.

92. Massidda O, Nováková L, Vollmer W. 2013. From models to pathogens: how much have we learned about *Streptococcus pneumoniae* cell division? Environmental Microbiology 15:3133–3157.

93. Straume D, Stamsås GA, Berg KH, Salehian Z, Håvarstein LS. 2017. Identification of pneumococcal proteins that are functionally linked to penicillin-binding protein 2b (PBP2b). Molecular Microbiology 103:99–116.

94. Tsui H-CT, Boersma MJ, Vella SA, Kocaoglu O, Kuru E, Peceny JK, Carlson EE, VanNieuwenhze MS, Brun YV, Shaw SL, Winkler ME. 2014. Pbp2x localizes separately from Pbp2b and other peptidoglycan synthesis proteins during later stages of cell division of *Streptococcus pneumoniae* D39. Molecular Microbiology 94:21–40.

95. Philippe J, Vernet T, Zapun A. 2014. The Elongation of Ovococci. Microb Drug Resist 20:215–221.

96. Atilano ML, Pereira PM, Yates J, Reed P, Veiga H, Pinho MG, Filipe SR. 2010. Teichoic acids are temporal and spatial regulators of peptidoglycan cross-linking in *Staphylococcus aureus*. Proceedings of the National Academy of Sciences 107:18991–18996.

97. Denapaite D, Brückner R, Hakenbeck R, Vollmer W. 2012. Biosynthesis of teichoic acids in *Streptococcus pneumoniae* and closely related species: lessons from genomes. Microb Drug Resist 18:344–358.

98. Fischer W. 1997. Pneumococcal Lipoteichoic and Teichoic Acid. Microbial Drug Resistance 3:309–325.

99. Bonnet J, Durmort C, Mortier-Barrière I, Campo N, Jacq M, Moriscot C, Straume D, Berg KH, Håvarstein L, Wong Y-S, Vernet T, Di Guilmi AM. 2018. Nascent teichoic acids insertion into the cell wall directs the localization and activity of the major pneumococcal autolysin LytA. The Cell Surface 2:24–37.

100. Kovács M, Halfmann A, Fedtke I, Heintz M, Peschel A, Vollmer W, Hakenbeck R, Brückner R. 2006. A Functional dlt Operon, Encoding Proteins Required for Incorporation of d-Alanine in Teichoic Acids in Gram-Positive Bacteria, Confers Resistance to Cationic Antimicrobial Peptides in *Streptococcus pneumoniae*. J Bacteriol 188:5797–5805.

101. Czabańska A, Neiwert O, Lindner B, Leigh J, Holst O, Duda KA. 2012. Structural analysis of the lipoteichoic acids isolated from bovine mastitis *Streptococcus uberis* 233, *Streptococcus dysgalactiae* 2023 and *Streptococcus agalactiae* 0250. Carbohydrate Research 361:200–205.

102. Zorzoli A, Meyer BH, Adair E, Torgov VI, Veselovsky VV, Danilov LL, Uhrin D, Dorfmueller HC. 2019. Group A, B, C, and G *Streptococcus* Lancefield antigen biosynthesis is initiated by a conserved α-d-GlcNAc-β-1,4-l-rhamnosyltransferase. Journal of Biological Chemistry 294:15237–15256.

103. Munoz E, Ghuysen J-M, Heymann H. 1967. Cell Walls of *Streptococcus pyogenes*, Type 14. C Polysaccharide-Peptidoglycan and G Polysaccharide-Peptidoglycan Complexes. Biochemistry 6:3659–3670.

104. Hols P, Hancy F, Fontaine L, Grossiord B, Prozzi D, Leblond-Bourget N, Decaris B, Bolotin A, Delorme C, Dusko Ehrlich S, Guédon E, Monnet V, Renault P, Kleerebezem M. 2005. New insights in the molecular biology and physiology of *Streptococcus thermophilus* revealed by comparative genomics. FEMS Microbiology Reviews 29:435–463.

105. Schultheiss AT. 2021. Overexpression, purification and characterization of the fratricin CbpD from *Streptococcus dysgalactiae*. Faculty of Chemistry, Biotechnology and Food Science, NMBU.

106. Lesk A. Introduction to genomics. Third edition. Oxford.

107. Invitrogen. 2008. pRSET A, B, and C: for high-level expression of recombinant proteins in *E.coli* Version F, Cat. no V351-20.

108. Simonian MH. 2002. Spectrophotometric Determination of Protein Concentration. Current Protocols in Food Analytical Chemistry 4:B1.3.1-B1.3.7.

109. SIB Swiss Institute of Bioinformatics. ExPASy - ProtParam tool. https://web.expasy.org/protparam/. Retrieved 24 April 2022.

110. Gill SC, von Hippel PH. 1989. Calculation of protein extinction coefficients from amino acid sequence data. Analytical Biochemistry 182:319–326.

111. Thermo Scientific. Extinction Coefficients: A guide to understanding extinction coefficients, with emphasis on spectrophotometric determination of protein concentration. Pierce Biotechnology https://doi.org/TR0006.4.

112. Clark DP, Pazdernik NJ, McGehee MR. 2019. Molecular Biology. Third edition. Elsevier.

113. Alberts B, Johnson A, Lewis J, Morgan D, Raff M, Roberts K, Walter P. 2015. Molecular Biology of The Cell. Sixth edition. W.W Norton & Company.

114. Kleppe K, Ohtsuka E, Kleppe R, Molineux I, Khorana HG. 1971. Studies on polynucleotides: XCVI. Repair replication of short synthetic DNA's as catalyzed by DNA polymerases. Journal of Molecular Biology 56:341–361.

115. Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA, Arnheim N. 1985. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science 230:1350–1354.

116. Delidow BC, Lynch JP, Peluso JJ, White BA. 1993. Polymerase Chain Reaction: Basic Protocols, p. 1–30. *In* PCR Protocols. Humana Press, New Jersey.

117. Rahman M, Uddin M, Sultana R, Moue A, Setu M. 2013. Polymerase Chain Reaction (PCR): A Short Review. Anwer Khan Modern Medical College Journal 4.

118. Campbell NA, Urry L, Cain M, Wassermann S, Minorsky P, Reece J. 2018. Biology - A Global Approach. Eleventh edition. Pearson, England.

119. Ramesh R, Munshi A, Panda SK. 1992. Polymerase chain reaction. Natl Med J India 5:115–119.

120. Rychlik W, Spencer WJ, Rhoads RE. 1990. Optimization of the annealing temperature for DNA amplification in vitro. Nucleic Acids Res 18:6409–6412.

121. New England Biolabs. https://international.neb.com. Phusion® High-Fidelity DNA Polymerase recombinant.

122. VWR Life Science. 2017. VWR Red Taq DNA Polymerase Master Mix. VWR International.

123. Bergkessel M, Guthrie C. 2013. Chapter Twenty Five - Colony PCR, p. 299–309. *In* Lorsch, J (ed.), Methods in Enzymology. Academic Press.

124. Hilgarth RS, Lanigan TM. 2019. Optimization of overlap extension PCR for efficient transgene construction. MethodsX 7:100759.

125. Zarghampoor F, Behzad-Behbahani A, Azarpira N, Khatami SR, Fanian M, Hossein Aghdaie M, Rafiei Dehbidi G. 2020. A Single Tube Overlap Extension PCR Method for Splicing of Multiple DNA Fragments. Avicenna J Med Biotechnol 12:37–43.

126. Serwer P. 1983. Agarose gels: Properties and use for electrophoresis. Electrophoresis 4:375–382.

127. Tiselius A. 1948. Electrophoresis and adsorption analysis as aids in investigations of large molecular weight substances and their breakdown products.

128. Barril P, Nates S. 2012. Introduction to Agarose and Polyacrylamide Gel Electrophoresis Matrices with Respect to Their Detection Sensitivities.

129. Lee PY, Costumbrado J, Hsu C-Y, Kim YH. 2012. Agarose Gel Electrophoresis for the Separation of DNA Fragments. J Vis Exp 3923.

130. DNA/RNA dye, peqGREEN. VWR. https://si.vwr.com. Retrieved 25 April 2022.

131. New England Biolabs. NEBuffer 2.1 | NEB. https://international.neb.com. Retrieved 9 February 2022.

132. New England Biolabs. NEBNext® Quick Ligation Reaction Buffer | NEB. https://international.neb.com. Retrieved 9 February 2022.

133. Liu I, Liu M, Shergill K. 2006. The Effect of Spheroplast Formation on the Transformation Efficiency in *Escherichia coli* DH5 9:5.

134. Sarkar S, Chaudhuri S, Basu T. 2002. Mechanism of artificial transformation of *E. coli* with plasmid DNA – Clues from the influence of ethanol. Current Science 83:1376–1380.

135. SOC Medium For use in transformation | Sigma-Aldrich. http://www.sigmaaldrich.com/. Retrieved 26 April 2022.

136. Eurofins Genomics. Eurofins Genomics - Genomic services by experts. https://eurofinsgenomics.eu/. Retrieved 10 February 2022.

137. Sanger F, Nicklen S, Coulson AR. 1977. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A 74:5463–5467.

138. Totomoch-Serra A, Marquez MF, Cervantes-Barragán DE. 2017. Sanger sequencing as a first-line approach for molecular diagnosis of Andersen-Tawil syndrome. F1000Res 6:1016.

139. Cloud-Based Informatics Platform for Life Sciences R&D. Benchling. https://www.benchling.com/. Retrieved 10 February 2022.

140. U.S National Library of Medicine. BLAST: Basic Local Alignment Search Tool. https://blast.ncbi.nlm.nih.gov. Retrieved 10 February 2022.

141. William Studier F, Rosenberg AH, Dunn JJ, Dubendorff JW. 1990. Use of T7 RNA polymerase to direct expression of cloned genes, p. 60–89. *In* Methods in Enzymology. Elsevier.

142. Kroll DJ, Abdel-Hafiz HA-M, Marcell T, Simpson S, Chen C-Y, Gutierrez-Hartmann A, Lustbader JW, Hoeffler JP. 1993. A Multifunctional Prokaryotic Protein Expression System: Overproduction, Affinity Purification, and Selective Detection. DNA and Cell Biology 12:441–453.

143. Spriestersbach A, Kubicek J, Schäfer F, Block H, Maertens B. 2015. Purification of His-Tagged Proteins, p. 1–15. *In* Methods in Enzymology. Elsevier.

144. Bornhorst JA, Falke JJ. 2000. [16] Purification of Proteins Using Polyhistidine Affinity Tags. Methods Enzymol 326:245–254.

145. Braich N, Codd R. 2008. Immobilised metal affinity chromatography for the capture of hydroxamate-containing siderophores and other Fe(iii)-binding metabolites directly from bacterial culture supernatants. Analyst 133:877.

146. Macherey-Nagel. 2019. Purification of His-tag proteins: User manual. Macherey-Nagel.

147. Manns JM. 2011. SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE) of Proteins. Current Protocols in Microbiology 22.

148. Walker JM. 1996. The Protein Protocols Handbook. Springer Science & Business Media.

149. Nowakowski AB, Wobig WJ, Petering DH. 2014. Native SDS-PAGE: High Resolution Electrophoretic Separation of Proteins With Retention of Native Properties Including Bound Metal Ions. Metallomics 6:1068–1078.

150. Lesk A. 2016. Introduction to protein science - architecture, function and genomics. Third Edition. Oxford University Press.

151. Vollmer W. Molecular biology of streptococci: Preparation and analysis of pneumococcal murein (peptidoglycan). Horizon Bioscience, Norfolk United Kingdom p 531-536.

152. Kotormán M, Laczkó I, Szabó A, Simon LM. 2003. Effects of Ca2+ on catalytic activity and conformation of trypsin and α-chymotrypsin in aqueous ethanol. Biochemical and Biophysical Research Communications 304:18–21.

153. Muga A, Arrondo JLR, Bellon T, Sancho J, Bernabeu C. 1993. Structural and Functional Studies on the Interaction of Sodium Dodecyl Sulfate with β-Galactosidase. Archives of Biochemistry and Biophysics 300:451–457.

154. Sanderson AR, Strominger JL, Nathenson SG. 1962. Chemical Structure of Teichoic Acid from *Staphylococcus aureus*, strain Copenhagen. Journal of Biological Chemistry 237:3603–3613.

155. Lambert T. Superfolder GFP at FPbase. FPbase. https://www.fpbase.org/protein/superfolder-gfp/. Retrieved 14 February 2022.

156. ImageJ. https://imagej.nih.gov/ij/. Retrieved 27 April 2022.

157. Ducret A. MicrobeJ – An ImageJ plug-in to analyze bacterial cells. https://www.microbej.com/. Retrieved 18 March 2022.

158. Ducret A, Quardokus EM, Brun YV. 2016. MicrobeJ, a tool for high throughput bacterial cell detection and quantitative analysis. 7. Nat Microbiol 1:1–7.

159. Beal J, Farny NG, Haddock-Angelli T, Selvarajah V, Baldwin GS, Buckley-Taylor R, Gershater M, Kiga D, Marken J, Sanchania V, Sison A, Workman CT. 2020. Robust estimation of bacterial cell count from optical density. Commun Biol 3:512.

160. Gaforio J j., Serrano M j., Ortega E, Algarra I, Alvarez de Cienfuegos G. 2002. Use of SYTOX green dye in the flow cytometric analysis of bacterial phagocytosis. Cytometry 48:93–96.

161. Boja ES, Fales HM. 2001. Overalkylation of a Protein Digest with Iodoacetamide. Anal Chem 73:3576–3582.

162. Majcherczyk PA, Rubli E, Heumann D, Glauser MP, Moreillon P. 2003. Teichoic Acids Are Not Required for *Streptococcus pneumoniae* and *Staphylococcus aureus* Cell Walls To Trigger the Release of Tumor Necrosis Factor by Peripheral Blood Monocytes. Infect Immun 71:3707–3713.

163. Romaniuk JAH, Cegelski L. 2018. Peptidoglycan and Teichoic Acid Levels and Alterations in *S. aureus* by Cell-Wall and Whole-Cell NMR. Biochemistry 57:3966–3975.

164. Eugster MR, Loessner MJ. 2011. Rapid Analysis of *Listeria monocytogenes* Cell Wall Teichoic Acid Carbohydrates by ESI-MS/MS. PLOS ONE 6:e21500.

165. EMBL-EBI. Clustal Omega < Multiple Sequence Alignment < EMBL-EBI. https://www.ebi.ac.uk. Retrieved 24 March 2022.

166. AlphaFold. https://www.deepmind.com/research/highlighted-research/alphafold. Retrieved 23 April 2022.

167. University of Washington. Robetta is a protein structure prediction service that is continually evaluated through CAMEO. https://robetta.bakerlab.org/. Retrieved 21 March 2022.

168. Schrödinger. PyMOL | pymol.org. https://pymol.org. Retrieved 21 March 2022.

169. Suzuki H, Lefébure T, Hubisz MJ, Pavinski Bitar P, Lang P, Siepel A, Stanhope MJ. 2011. Comparative Genomic Analysis of the *Streptococcus dysgalactiae* Species Group: Gene Content, Molecular Adaptation, and Promoter Evolution. Genome Biol Evol 3:168–185.

170. Huber R. 1987. Flexibility and rigidity, requirements for the function of proteins and protein pigment complexes. Eleventh Keilin memorial lecture. Biochem Soc Trans 15:1009–1020.

171. Gregersen N, Bross P, Vang S, Christensen JH. 2006. Protein Misfolding and Human Disease. Annual Review of Genomics and Human Genetics 7:103–124.

172. Biørnstad TJ, Ohnstad HS, Håvarstein LS 2012. Deletion of the murein hydrolase CbpD reduces transformation efficiency in *Streptococcus thermophilus*. Microbiology 158:877–885.

173. Clemans DL, Kolenbrander PE, Debabov DV, Zhang Q, Lunsford RD, Sakone H, Whittaker CJ, Heaton MP, Neuhaus FC. 1999. Insertional Inactivation of Genes Responsible for the d-Alanylation of Lipoteichoic Acid in *Streptococcus gordonii* DL1 (Challis) Affects Intrageneric Coaggregations. Infect Immun 67:2464–2474.

174. Perego M, Glaser P, Minutello A, Strauch MA, Leopold K, Fischer W. 1995. Incorporation of D-Alanine into Lipoteichoic Acid and Wall Teichoic Acid in *Bacillus subtilis*: identification of genes and regulation. Journal of Biological Chemistry 270:15598–15606.

175. Miyauchi E, Morita M, Rossi M, Morita H, Suzuki T, Tanabe S. 2012. Effect of D-alanine in teichoic acid from the *Streptococcus thermophilus* cell wall on the barrier-protection of intestinal epithelial cells. Biosci Biotechnol Biochem 76:283–288.

176. Marfey P. 1984. Determination of D-amino acids. II. Use of a bifunctional reagent, 1,5-difluoro-2,4-dinitrobenzene. Carlsberg Research Communications 49:591–596.

177. Brown S, Santa Maria JP, Walker S. 2013. Wall Teichoic Acids of Gram-Positive Bacteria. Annu Rev Microbiol 67:10.1146/annurev-micro-092412–155620.

178. McCormick NE, Halperin SA, Song F., LeeYR 2011. Regulation of d-alanylation of lipoteichoic acid in *Streptococcus gordonii*. Microbiology 157:2248–2256.

179. Koprivnjak T, Mlakar V, Swanson L, Fournier B, Peschel A, Weiss JP. 2006. Cation-Induced Transcriptional Regulation of the dlt Operon of *Staphylococcus aureus*. Journal of Bacteriology 188:3622–3630.

180. Bartual SG, Straume D, Stamsås GA, Muñoz IG, Alfonso C, Martínez-Ripoll M, Håvarstein LS, Hermoso JA. 2014. Structural basis of PcsB-mediated cell separation in *Streptococcus pneumoniae*. 1. Nat Commun 5:3842.

181. Bailey TL, Johnson J, Grant CE, Noble WS. 2015. The MEME Suite. Nucleic Acids Research 43:W39–W49.

182. Hakenbeck R, Brückner R, Denapaite D, Maurer P. 2012. Molecular mechanisms of β-lactam resistance in *Streptococcus pneumoniae*. Future Microbiology 7:395–410.

183. Zapun A, Contreras-Martel C, Vernet T. 2008. Penicillin-binding proteins and β-lactam resistance. FEMS Microbiol Rev 32:361–385.

184. Chambers HF. 1999. Penicillin-Binding Protein-Mediated Resistance in Pneumococci and Staphylococci. The Journal of Infectious Diseases 179:S353–S359.

185. Beres SB, Zhu L, Pruitt L, Olsen RJ, Faili A, Kayal S, Musser JM. Integrative Reverse Genetic Analysis Identifies Polymorphisms Contributing to Decreased Antimicrobial Agent Susceptibility in *Streptococcus pyogenes*. mBio 13:e03618-21.

186. Jordal S, Glambek M, Oppegaard O, Kittang BR. 2015. New Tricks from an Old Cow: Infective Endocarditis Caused by *Streptococcus dysgalactiae* subsp. *dysgalactiae.* J Clin Microbiol 53:731–734.

187. Park MJ, Eun I-S, Jung C-Y, Ko Y-C, Kim Y-J, Kim C, Kang E-J. 2012. *Streptococcus dysgalactiae* subspecies *dysgalactiae* Infection after Total Knee Arthroplasty: A Case Report. Knee Surg Relat Res 24:120–123.

Appendix 1: Preparation of C-medium

C-medium was used as medium for growth curve assay with *Streptococcus dysgalactiae* (sections 4.4 and 4.5). This was done as the $CbpD^{SD}$ protein works best under the conditions provided by this medium. In order to make C-medium (section 2.9.1), a pre-C-medium, ADAMS I, ADAMS II and ADAMS III solutions were necessary and had to be made in advance. The following recipes for the various solution are given:

ADAMS I

0.15 mL 0.5 mg/mL biotin, 7.5 mg niacin, 87.5 mg pyridoxine hydrochloride, 300 mg calcium pantothenate, 800 mg thiamine hydrochloride, 35 mg riboflavin

Adjusted to 0.5 L with dH_2O, with pH adjusted to 7.0. Solution was sterile filtered, and stored at 4 ℃.

ADAMS II (10x)

500 mg iron sulfate heptahydrate, 500 mg copper sulfate pentahydrate, 500 mg zinc sulfate heptahydrate, 200 mg manganese chloride tetrahydrate, 10 mL concentrated HCl

Adjusted to 100 mL with dH_2O, sterile filtered and stored at 4 ℃.

ADAMS III

128 mL ADAMS I, 3.2 mL ADAMS II (10x), 1.6 mL asparagine monohydrate, 0.16 g choline chloride, 0.4 g calcium chloride, 16 g magnesium chloride hexahydrate

Adjusted to 800 mL with dH_2O, with pH adjusted to 7.6. Solution was sterile filtered, and stored at 4 ℃.

Pre-C medium

0.045 g cysteine hydrochloride, 8 g sodium acetate, 20 g Bacto™ casitone, 0.024 g L-tryptophan, 34 g dipotassium hydrogen phosphate

Adjusted to 4 L with dH_2O, and autoclaved.

Appendix 2: CbpD^{SD} and sfGFP-CbpD^{SD} amino acid sequences

The CbpD^{SD} and sfGFP-CbpD^{SD} amino acid sequences are given below, where each domain is colored accordingly. The signal sequence for a mature CbpD^{SD} is also included. The CbpD^{SD} sequence (1131 bp) contains an underlined leader peptide, the CHAP domain highlighted in blue, a linker between CHAP and SH3b domains, the SH3b domain highlighted in yellow and lastly the conserved domain in black. Similarly, the sfGFP-CbpD^{SD} (1371 bp) contains the same linker, SH3b domain in yellow and the conserved domain in black. The CHAP domain has been replaced by a sfGFP sequence, highlighted in green which also include the (poly)His-tag.

>CbpD dysgalactiae with signal sequence

MKKIHQLLVSGAILLSVNGAVSSVASTLNAEHTGVVHAAVLGDNYPSKWKKGSGIDSWNMYVRQCTSFVAFRLSSANGFQL
PKGYGNACTWGHIAKKQGYTVNKTPKVGAVAWFDTNAFQSHATYGHVAWVAEVRGDSVVIEEYNYNAGQGPEKYHKRQ
IPKNHVSGYIHFKDLPSGEASKSQTKEQQVSKEAVKQGGTYHFTERTPVKAQAQLTSPDLAYYNPGQSVHYDQAMTVD
GHEWISYLSFSGSRRYIPIKKTGQKTQQVSETTSPINIGDRVTFPGVFRVDRIVNNLLVSEELAGGGATSLNWIDPSP
LDETDRKGVKAGNQILQAGEFFVIPGNYRVLKVDRPSNGIYVKIGSRGTWLTADKASKLP-

>sfGFP-CbpD dysgalactiae

MHHHHHHKHLTGSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTLTYG
VQCFSRYPDHMKQHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNF
NSHNVYITADKQKNGIKANFKIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSVLSKDPNEKRDHMVLLEF
VTAAGITHGMDELYKTSGAAAPSGEASKSQTKEQQVSKEAVKQGGTYHFTERTPVKAQAQLTSPDLAYYNPGQSVHYD
QAMTVDGHEWISYLSFSGSRRYIPIKKTGQKTQQVSETTSPINIGDRVTFPGVFRVDRIVNNLLVSEELAGGGATSLN
WIDPSPLDETDRKGVKAGNQILQAGEFFVIPGNYRVLKVDRPSNGIYVKIGSRGTWLTADKASKLP-

Appendix 3: Multiple sequence alignment of CbpDs from various streptococcal species

The complete multiple sequence alignment produced with query protein sequence CbpD^{SD} (without leader peptide) against streptococcal species *S. thermophilus* (AKH33394.1**)**, *S. vestibularis* (WP_117648089), *S. porcinus* (WP_003083717), *S. uberis* (WP_203261572.1), *S. iniae* (WP_071127360), *S.* (WP_008089638.1), *S. equi* (WP_012677198.1), *S. pyogenes* (WP_136094586.1) CbpDs sequences. The MSA includes the CHAP domain, the SH3b domain and the conserved domain, although not all species possess an SH3b domain in their fratricin. Furthermore, the MSA reveals the conserved cysteine (C) of the CHAP domain, the LAGG (leucine-alanine-glycine-glycine) motif and the conserved amino acids glutamine (Q) and tryptophan (W).

```
AKH33394.1          MEVEMKKKLLNYLILL-------ACLATVSVSCSGSAQAAVLGDDYPFSWKYGGFGVDPW      53
WP_117648089        ----MKKKLLNCLLPL-------ACLATISVSCGSSAQAAVLGDDYPSSWKYGGFGVDPW      49
WP_003083717        MK-YFTNKGISFLL------FVSLGLATLFPSCCQPVTAAVIGDDYPINWKS-GLGSDTW      52
WP_203261572.1      MK-FIKILLS-QIVSL-------FLLLTISLHSLETVNAAVLGDDYPIAWKS-GWGTDTW      50
WP_071127360        MK-ILKQVLSLTLVLTSLWGIQTLPLTNKSLAFSKTVKAAVLGDNYPHKWQK-GWGADSW      58
WP_008089638.1      MK-KFHQLLVSGAILLSVNGAISTAAPLFNPNSVSQVQAEVLGDNYPSKWKK-ESGIDSW      58
WP_012677198.1      MK-KVYQLLTSGAILLGVNGAISSTTTAISQGQSGIVHAAVLGDNYPSKWKQ-GFGADPW      58
WP_136094586.1      MK-KFHRFLVSGVILLGFNGLVPTMPSTLISQHENLVHAAVLGDNYPSKWKK-GNGIDSW      58
CbpD                ------------------------------EHTGVVHAAVLGDNYPSKWKK-GSGIDSW      28
                                                   . * *:**:**  *:    * * *

AKH33394.1          TMYWRQCTSFAAYRLSNTNGFTLPIGYGNAITWGNIARANGYRVDMNPAVGSIAWFSAGV      113
WP_117648089        TMYWRQCTSFAAYRLSNSNGFTLPVGYGNAITWGPIARANGHRVDMNPAVGSIAWFSNGV      109
WP_003083717        GMYLRQCTSFVAYRLHIVNGFSLPSGYGNADSWGHVARRNGYQVDINPQVGSVAWFDKGV      112
WP_203261572.1      GMYRRQCTSFVAFRLNQVNGFSIPSGLGNADTWGHIARRMGYPVNDIPAVGAVAWFDKGI      110
WP_071127360        GMYLRQCTSFVAFRLSTANGFNLPAGYGNADSWGHIAKAQGFLVNQIPQVGAVAWFDKGV      118
WP_008089638.1      NMYVRQCTSFVAFRLSSANGFQLPRGYGNAHTWGHIAKTQGYAVNKTPKVGSVAWFDKGA      118
WP_012677198.1      NMYLRQCTSFVAFRLHSANGFSLPRGYGNAESWGHIAKHQGYLVDHTPRVGAVAWWDKGF      118
WP_136094586.1      NMYIRQCTSFAAFRLSSANGFQLPKGYGNACTWGHIAKNQGYPVNKTPSIGAIAWFDKNA      118
CbpD                NMYVRQCTSFVAFRLSSANGFQLPKGYGNACTWGHIAKKQGYTVNKTPKVGAVAWFDTNA      88
                     ** ******.*:**   *** :* * *** :** :*:  *. *:  * :*::**:. .

AKH33394.1          NGA-GHMGHVAWVAEVHGDQVTIEEYNYDAGQGPEKYHKRSFHKSQVSGYIHFKDLESGT      172
WP_117648089        NGA-GYMGHVAWVAEVNGDQVTIEEYNYNAGQGPEKYHKRSFHKSQVSGYIHFKDLESGT      168
WP_003083717        SLSHREYGHVAWVAEVKGDLVTLEEYNYNAGQGPEKYHRRQILRQEASGYIHFKNLINNS      172
WP_203261572.1      NGSHLAYGHVSWVAEVNGNLVTLEEYNFDAGQGPEQYHRRVINSHQVSGFIHFKDVDTGS      170
WP_071127360        NYSHRDYGHVAWVAEVSGDYVTLEEYNYNAGQGPEKYHRRQIHRQQVSGFIHFKDITTEQ      178
WP_008089638.1      NNSSAIYGHVSWVAEVSNDYVTLEEYNYNAGQGPEKYHKRQIHKSKVSGFIHFKDLPSTT      178
WP_012677198.1      NQSHALYGHVAWVAEVNGDTVIIEEYNYNAGQGPEKYHKRQIHKHQVSGYIHFKDLDHNS      178
WP_136094586.1      YQSNAAYGHVAWVADIRGDTVTIEEYNYNAGQGPERYHKRQIPKSQVSGYIHFKDLSSQT      178
CbpD                FQSHATYGHVAWVAEVRGDSVVIEEYNYNAGQGPEKYHKRQIPKNHVSGYIHFKDLPSGE      148
                      :    ***:***:: .: * :****::******:**:* :   ..**:****::

AKH33394.1          QNRNP-------------------------------------------------------      177
WP_117648089        QNGNP-------------------------------------------------------      173
WP_003083717        SNISI----SQLSRHSYLAPKGSYTFNQRKAVKAQPSLNSSDLVYYEKGNTVHYDKILDA      228
WP_203261572.1      IPISS----PMQSPRSSIPNQGSYHFTEQMPIKAQPIVNSQAIAFYQAGQMVHYDKTIIA      226
WP_071127360        STIPS-SKHILAESTKLLPNKGTYYFKAMSPVKEEARFQSLTLTSYSAGQYVHYDQTMVA      237
WP_008089638.1      SSSSVEINQPLSNKDKTLPKSGLYQFTQKHLIKGEAKIDSPDLGYYDAGQSVNYDQLLSS      238
WP_012677198.1      TSHTQML-HASQAGPARLAKSGTYYFSSQSPIKAEATLASPDLAYYHTGQLVHYDQTLVA      237
WP_136094586.1      SHSYP-R-QLKHISQSSFDPSGTYHFTTRLPVKGQTSIDSPDLAYYEAGQSVYYDKVVTA      236
CbpD                ASKSQ-T-KEQQVSKEAVKQGGTYHFTERTPVKAQAQLTSPDLAYYNPGQSVHYDQAMTV      206
```

Appendix 3: Multiple sequence alignment of CbpDs from various streptococcal species

```
AKH33394.1          ----------------------------------------TNSSIKVGDSVRFTGTFR      195
WP_117648089        ----------------------------------------TNSSIKVGDSVRFTGTFR      191
WP_003083717        DGYRWLSYIGGSGNRRYIAIEKLEHQV---------------ETNDFKVEQIVHFKGPFQ      273
WP_203261572.1      DGYQWLSYIAYSGQRRYIPIAKVTSKESS-------------KQEDFAPGDQVTFSGVYQ      273
WP_071127360        DGYQWISYIGFSGNRRYIPVVKIEERKSES-------PETPKSERIFKVGDTVSFQGTFK      290
WP_008089638.1      DGYDWLSYLSFSGNRRYIPLKKITEKSNTVEVTHAKTTTTTAIQNPIVVGDQVTFPGVFK      298
WP_012677198.1      DGYEWLSYIGHSGNRRYIPIHKLPVQPQQHSQGA-IDKPKPVATTPIKVGDTVAFPGVFR      296
WP_136094586.1      GGYTWLSYLSFSGNRRYIPIKEPAQSVVQN----------DNTKPSIKVGDTVTFPGVFR      286
CbpD                DGHEWISYLSFSGSRRYIPIKKTGQKTQQV----------SETTSPINIGDRVTFPGVFR      256
                                                                :   : * * * ::

AKH33394.1          VTSVSGNTITSQDLAGGTPTKHNIVDPSPVLEVDAQGNPTSDQYLNPGETFTIPGNFKVL      255
WP_117648089        VTSVSGNTITSQDLAGGTPTKHNIVDPGPVLEVDGQGNPTSDQYLNPGETFTIPGNFKVL      251
WP_003083717        VSHNHGWMILSNILAGGAATQLNWLDPGPLIETDKIGQKAGDQVLYPGEYFSIPGEFKIL      333
WP_203261572.1      VTQIHGSLLSSKDLAGGEPGPLNWLDPGPILESNRDGQQSGDQILYPGDFFIIPGNYKVL      333
WP_071127360        VTQIHGNLVSSADLAGGQPSQLNWIDPGPVDKTNASGQKEANQILSPDNYFTIPGTYKVL      350
WP_008089638.1      VTQVINGLITSSQLAGGDPTHYNWIDPGPLTETDAKGNTIGDQILHSGEYFTVPGTYKVL      358
WP_012677198.1      VDQVAQNMIASSELAGGEPTSLNWIDPAPLHETDHKGRIAGNQILQVGDYFVISGTYKVL      356
WP_136094586.1      VDQLVNNLIVNKELAGGDPTPLNWIDPTPLDETDNQGKVLGDQILRVGEYFIVTGSYKVL      346
CbpD                VDRIVNNLLVSEELAGGGATSLNWIDPSPLDETDRKGVKAGNQILQAGEFFVIPGNYRVL      316
                    *       : .  ****     * :** *: : :  *   .:* *  .: * : * :::*

AKH33394.1          AIDPPSDGILVQIGNLKTWVTQSVLEKI-        283
WP_117648089        AIDPPSDGVLVQIGNRKTWVTQSVLEKA-        279
WP_003083717        QVDQESKGIKIKIGQKQTWISMDQVHKK-        361
WP_203261572.1      QIHKETKGLLIQIGNRQTWVSMNKVQKST        362
WP_071127360        QVDLPSRGLYIQIGSRASWVTMDRVLKQ-        378
WP_008089638.1      KVDKPSNGIYVQMGSRGTWFIMDKAKKV-        386
WP_012677198.1      KVDQPSKGIYVQMGSRGTWLTIDKAKKV-        384
WP_136094586.1      KIDQPSNGIYVQIGSRGTWVNADKANKL-        374
CbpD                KVDRPSNGIYVKIGSRGTWLTADKASKLP        345
                     :.  : *: :::*.  :*.  .   *
```

Figure A.3: Complete multiple sequence alignment with CbpD^{SD} as query against streptococcal species *S. thermophilus* (AKH33394.1**)**, *S. vestibularis* (WP_117648089), *S. porcinus* (WP_003083717), *S. uberis* (WP_203261572.1), *S. iniae* (WP_071127360), *S.* (WP_008089638.1), *S. equi* (WP_012677198.1), *S. pyogenes* (WP_136094586.1) CbpD protein sequences.

Appendix 4: CbpD prediction with RoseTTAFold

Protein structure prediction service RoseTTAFold (167) was used to make a prediction of the $CbpD^{SD}$ protein by entering the MA201 amino acid sequence without its signal sequence. The prediction with lowest angstroms error estimates over positions 0-300 was fetched as a PDB-file and uploaded to PyMOL (168) in order to visualize the protein as a cartoon. The predicated 3D structure of $CbpD^{SD}$ shows a typical CHAP domain with an open cavity where the catalytical cysteine is located. The CHAP is connected to the SH3b domain via an α-helix. which perhaps gives flexibility to the SH3b domain to bind to the peptidoglycan. Furthermore, the SH3b domain are predicted to be built up by seven β-strands that form two β-sheets packed against each other forming a typical β-barrel fold, while the conserved domain is predicted to consist of eight anti parallel β-strands that form a cylindric like structure.

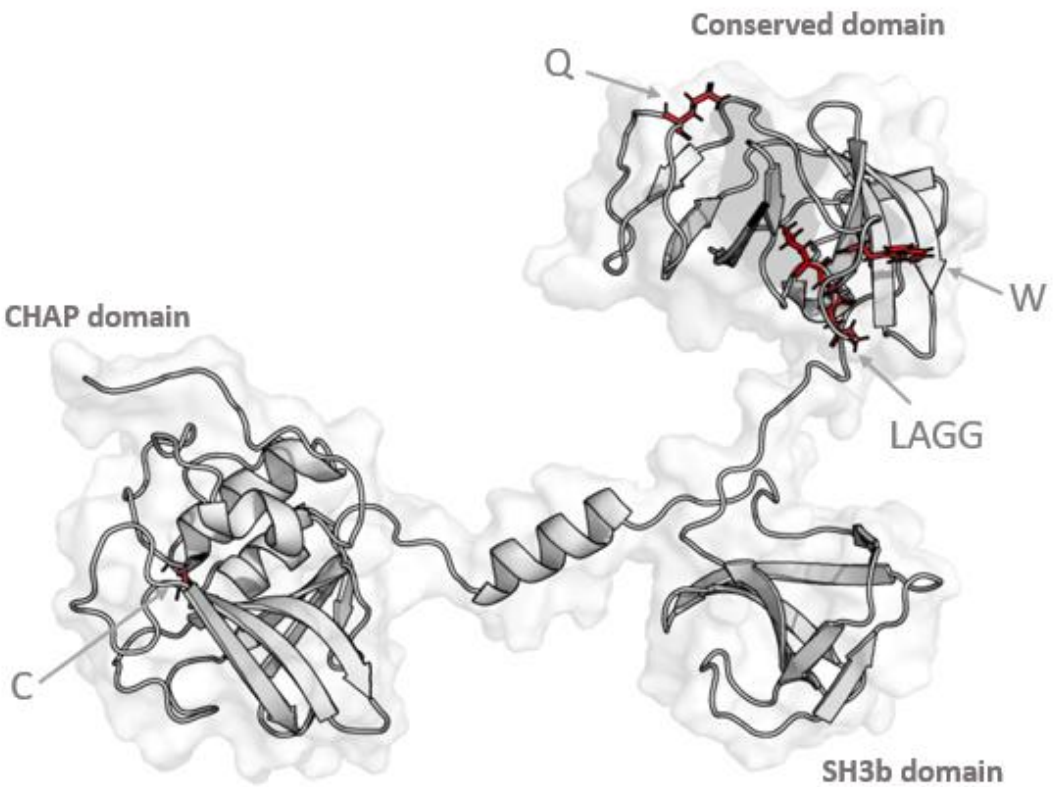

Figure A.4: RoseTTAFold predicted 3D structure of $CbpD^{SD}$ of *S. dysgalactiae* strain MA201. The various domains are named (CHAP, SH3b or conserved domain) with grey arrows pointing to conserved amino acids in the sequence. The cysteine (C) of the CHAP domain, glutamine (Q), tryptophan (W) and the leucine-alanine-glycine-glycine motif (LAGG) are the important conserved amino acids in this work and presented as red sticks. Meanwhile, the predicted structure is presented as a cartoon with β-strands, α-helixes and loops displayed, with a transparent surface behind. Created with PyMOL (168).

Appendix 5: Point-mutated sfGFP-CbpD$^{\Delta SH3b}$ sequences with site-directed mutations

The DNA sequences below show the complete sequence with the site-directed mutations necessary to create the recombinant proteins: a sfGFP-CbpD$^{\Delta SH3b,L300S,A301S}$, sfGFP-CbpD$^{\Delta SH3b,G302A,G303A}$, sfGFP-CbpD$^{\Delta SH3b,Q329A}$ and a sfGFP-CbpD$^{\Delta SH3b,W365A}$. The nucleotide bases, which have been changed into the desired amino acids (Ala or Ser), are displayed in yellow.

> LAGG to SSGG

ATGCATCATCATCATCATCATAAACATCTTACCGGTTCTAAAGGTGAAGAGTTGTTTACAGGTGTTGTTCC
AATCTTGGTTGAGTTGGATGGTGATGTTAATGGGCATAAGTTTTCTGTTCGTGGTGAAGGTGAAGGTGATG
CTACAAATGGTAAACTCACCCTCAAGTTTATTTGTACCACAGGTAAATTGCCAGTTCCATGGCCAACATTG
GTTACAACATTGACTTATGGTGTTCAATGTTTTTCTCGTTACCCAGATCACATGAAGCAGCACGATTTTTT
CAAATCTGCTATGCCAGAAGGTTACGTTCAAGAACGTACCATCTCTTTTAAGGATGATGGAACCTATAAAA
CCCGTGCTGAAGTTAAGTTTGAAGGTGATACATTGGTTAATCGTATTGAATTGAAAGGTATTGATTTTAAA
GAAGATGGTAATATTCTTGGTCATAAATTGGAATATAATTTTAATTCTCATAATGTTTATATCACAGCTGA
TAAACAAAAAAAATGGTATTAAAGCTAATTTCAAAATCCGTCATAATGTTGAAGATGGTTCTGTCCAATTGG
CTGATCATTATCAACAGAATACCCCAATTGGTGATGGTCCAGTTTTGTTGCCAGATAATCATTATTTGTCT
ACACAAAGTGTTTTGTCTAAAGATCCAAATGAGAAGCGTGATCACATGGTCTTGTTGGAATTTGTTACAGC
TGCTGGTATTACCCATGGTATGGATGAATTGTACAAAACTAGTGGAGCGGCCGCACCTAGTGGTGAAGCAA
GTAAGTCTCAAACAAAAGAACAACAGGTTTCTAAAGAAGCAGTAAAAAAGACGGGGCAAAAAACACAACAA
GTCTCTGAGACAACATCGCCTATCAATATTGGTGATAGAGTGACTTTCCCTGGCGTTTTCCGTGTGGATCG
TATTGTAAACAATCTATTAGTTAGCGAGGAG**AGTAGT**GGTGGGGGCGCTACTTCTCTTAACTGGATTGATC
CCTCACCTTTGGATGAAACAGATCGTAAAGGAGTAAAAGCAGGAAATCAAATTTTACAGGCCGGTGAGTTT
TTTGTTATCCCAGGTAACTATAGGGTACTGAAAGTCGACCGACCAAGCAATGGGATTTATGTCAAGATTGG
ATCACGTGGAACATGGTTAACTGCTGATAAAGCCTCTAAATTGCCATAAAAGCTT-

> LAGG to LAAA

ATGCATCATCATCATCATCATAAACATCTTACCGGTTCTAAAGGTGAAGAGTTGTTTACAGGTGTTGTTCC
AATCTTGGTTGAGTTGGATGGTGATGTTAATGGGCATAAGTTTTCTGTTCGTGGTGAAGGTGAAGGTGATG
CTACAAATGGTAAACTCACCCTCAAGTTTATTTGTACCACAGGTAAATTGCCAGTTCCATGGCCAACATTG
GTTACAACATTGACTTATGGTGTTCAATGTTTTTCTCGTTACCCAGATCACATGAAGCAGCACGATTTTTT
CAAATCTGCTATGCCAGAAGGTTACGTTCAAGAACGTACCATCTCTTTTAAGGATGATGGAACCTATAAAA
CCCGTGCTGAAGTTAAGTTTGAAGGTGATACATTGGTTAATCGTATTGAATTGAAAGGTATTGATTTTAAA
GAAGATGGTAATATTCTTGGTCATAAATTGGAATATAATTTTAATTCTCATAATGTTTATATCACAGCTGA
TAAACAAAAAAAATGGTATTAAAGCTAATTTCAAAATCCGTCATAATGTTGAAGATGGTTCTGTCCAATTGG
CTGATCATTATCAACAGAATACCCCAATTGGTGATGGTCCAGTTTTGTTGCCAGATAATCATTATTTGTCT
ACACAAAGTGTTTTGTCTAAAGATCCAAATGAGAAGCGTGATCACATGGTCTTGTTGGAATTTGTTACAGC
TGCTGGTATTACCCATGGTATGGATGAATTGTACAAAACTAGTGGAGCGGCCGCACCTAGTGGTGAAGCAA
GTAAGTCTCAAACAAAAGAACAACAGGTTTCTAAAGAAGCAGTAAAAAAGACGGGGCAAAAAACACAACAA
GTCTCTGAGACAACATCGCCTATCAATATTGGTGATAGAGTGACTTTCCCTGGCGTTTTCCGTGTGGATCG
TATTGTAAACAATCTATTAGTTAGCGAGGAGCTAGCT**GCTGCT**GGCGCTACTTCTCTTAACTGGATTGATC
CCTCACCTTTGGATGAAACAGATCGTAAAGGAGTAAAAGCAGGAAATCAAATTTTACAGGCCGGTGAGTTT
TTTGTTATCCCAGGTAACTATAGGGTACTGAAAGTCGACCGACCAAGCAATGGGATTTATGTCAAGATTGG
ATCACGTGGAACATGGTTAACTGCTGATAAAGCCTCTAAATTGCCATAAAAGCTT-

Appendix 5: Point-mutated sfGFP-CbpD□SH3b sequences with site-directed mutations

> Q to A

ATGCATCATCATCATCATCATAAACATCTTACCGGTTCTAAAGGTGAAGAGTTGTTTACAGGTGTTGTTCC
AATCTTGGTTGAGTTGGATGGTGATGTTAATGGGCATAAGTTTTCTGTTCGTGGTGAAGGTGAAGGTGATG
CTACAAATGGTAAACTCACCCTCAAGTTTATTTGTACCACAGGTAAATTGCCAGTTCCATGGCCAACATTG
GTTACAACATTGACTTATGGTGTTCAATGTTTTTCTCGTTACCCAGATCACATGAAGCAGCACGATTTTTT
CAAATCTGCTATGCCAGAAGGTTACGTTCAAGAACGTACCATCTCTTTTAAGGATGATGGAACCTATAAAA
CCCGTGCTGAAGTTAAGTTTGAAGGTGATACATTGGTTAATCGTATTGAATTGAAAGGTATTGATTTTAAA
GAAGATGGTAATATTCTTGGTCATAAATTGGAATATAATTTTAATTCTCATAATGTTTATATCACAGCTGA
TAAACAAAAAAAATGGTATTAAAGCTAATTTCAAAATCCGTCATAATGTTGAAGATGGTTCTGTCCAATTGG
CTGATCATTATCAACAGAATACCCCAATTGGTGATGGTCCAGTTTTGTTGCCAGATAATCATTATTTGTCT
ACACAAAGTGTTTTGTCTAAAGATCCAAATGAGAAGCGTGATCACATGGTCTTGTTGGAATTTGTTACAGC
TGCTGGTATTACCCATGGTATGGATGAATTGTACAAAACTAGTGGAGCGGCCGCACCTAGTGGTGAAGCAA
GTAAGTCTCAAACAAAAGAACAACAGGTTTCTAAAGAAGCAGTAAAAAAGACGGGGCAAAAAACACAACAA
GTCTCTGAGACAACATCGCCTATCAATATTGGTGATAGAGTGACTTTCCCTGGCGTTTTCCGTGTGGATCG
TATTGTAAACAATCTATTAGTTAGCGAGGAGCTAGCTGGTGGGGGCGCTACTTCTCTTAACTGGATTGATC
CCTCACCTTTGGATGAAACAGATCGTAAAGGAGTAAAAGCAGGAAAT**GCT**ATTTTACAGGCCGGTGAGTTT
TTTGTTATCCCAGGTAACTATAGGGTACTGAAAGTCGACCGACCAAGCAATGGGATTTATGTCAAGATTGG
ATCACGTGGAACATGGTTAACTGCTGATAAAGCCTCTAAATTGCCATAAAAGCTT-

> W to A

ATGCATCATCATCATCATCATAAACATCTTACCGGTTCTAAAGGTGAAGAGTTGTTTACAGGTGTTGTTCC
AATCTTGGTTGAGTTGGATGGTGATGTTAATGGGCATAAGTTTTCTGTTCGTGGTGAAGGTGAAGGTGATG
CTACAAATGGTAAACTCACCCTCAAGTTTATTTGTACCACAGGTAAATTGCCAGTTCCATGGCCAACATTG
GTTACAACATTGACTTATGGTGTTCAATGTTTTTCTCGTTACCCAGATCACATGAAGCAGCACGATTTTTT
CAAATCTGCTATGCCAGAAGGTTACGTTCAAGAACGTACCATCTCTTTTAAGGATGATGGAACCTATAAAA
CCCGTGCTGAAGTTAAGTTTGAAGGTGATACATTGGTTAATCGTATTGAATTGAAAGGTATTGATTTTAAA
GAAGATGGTAATATTCTTGGTCATAAATTGGAATATAATTTTAATTCTCATAATGTTTATATCACAGCTGA
TAAACAAAAAAAATGGTATTAAAGCTAATTTCAAAATCCGTCATAATGTTGAAGATGGTTCTGTCCAATTGG
CTGATCATTATCAACAGAATACCCCAATTGGTGATGGTCCAGTTTTGTTGCCAGATAATCATTATTTGTCT
ACACAAAGTGTTTTGTCTAAAGATCCAAATGAGAAGCGTGATCACATGGTCTTGTTGGAATTTGTTACAGC
TGCTGGTATTACCCATGGTATGGATGAATTGTACAAAACTAGTGGAGCGGCCGCACCTAGTGGTGAAGCAA
GTAAGTCTCAAACAAAAGAACAACAGGTTTCTAAAGAAGCAGTAAAAAAGACGGGGCAAAAAACACAACAA
GTCTCTGAGACAACATCGCCTATCAATATTGGTGATAGAGTGACTTTCCCTGGCGTTTTCCGTGTGGATCG
TATTGTAAACAATCTATTAGTTAGCGAGGAGCTAGCTGGTGGGGGCGCTACTTCTCTTAACTGGATTGATC
CCTCACCTTTGGATGAAACAGATCGTAAAGGAGTAAAAGCAGGAAATCAAATTTTACAGGCCGGTGAGTTT
TTTGTTATCCCAGGTAACTATAGGGTACTGAAAGTCGACCGACCAAGCAATGGGATTTATGTCAAGATTGG
ATCACGTGGAACA**GCT**TTAACTGCTGATAAAGCCTCTAAATTGCCATAAAAGCTT-

Appendix 6: CbpD^{SD} and CbpD^{C65A} for growth curve assay

The following Figures A.6.1 and A.6.2 displays the elution and purification of CbpD^{SD} wild-type (A and B) and CbpD^{C65A} (C and D) respectively. The proteins were eluted from 400 mL cultures, with a high 280 nm absorbance peak registered after 10 mL of linear gradient elution for CbpD^{SD} and low absorbance peak after 3 mL for CbpD^{C65A}. Fractions 10-13 were therefore collected for CbpD^{SD}, which collectively had a protein yield of 2.55 mg/mL (NanoDrop™ 2000). Fractions 3-6 were collected for CbpD^{C65A}. Based on the SDS-PAGE and Coomassie blue staining, the purity of recombinant protein was in the range of ~80% for CbpD^{SD} while in the range of ~75% for CbpD^{C65A}. Additionally, the CbpD^{SD} as seen on the SDS-PAGE has a band size of ~43 kilodaltons (kDa), which is in accordance with its theoretical molecular weight of 40 kDa (Figure A.6.1B). For the CbpD^{C65A} protein, overexpression by IPTG did not work as indicated by the elution graph and the SDS-PAGE, and it was believed that the protein was not transcribed at all (Figure A.6.2D).

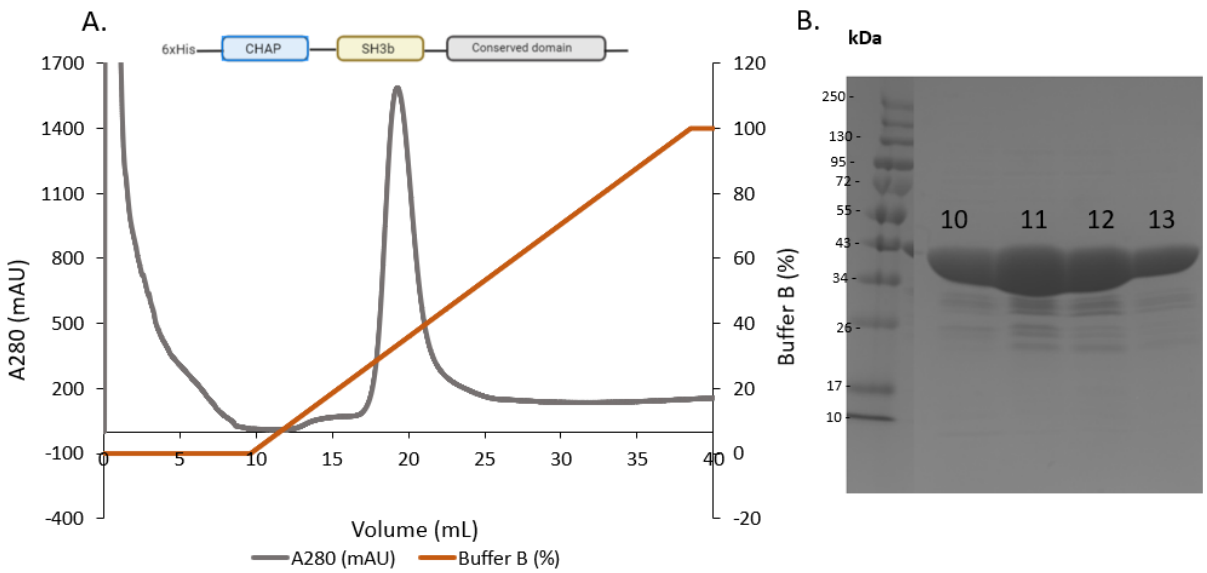

Figure A.6.1: A. Purification of CbpD^{SD} with the graph displaying 280 nm absorbance of the protein (grey line) after about 10 mL of linear gradient elution with Buffer B (500 mM free imidazole) (orange line). Fractions 10-13 were therefore collected. **B.** SDS-PAGE confirms the presence of CbpD^{SD} as correct band sizes (40.04 kDa) is displayed. The purity of the recombinant protein was believed to be in the range of ~80%.

Appendix 6: CbpDSD and CbpDC65A for growth curve assay

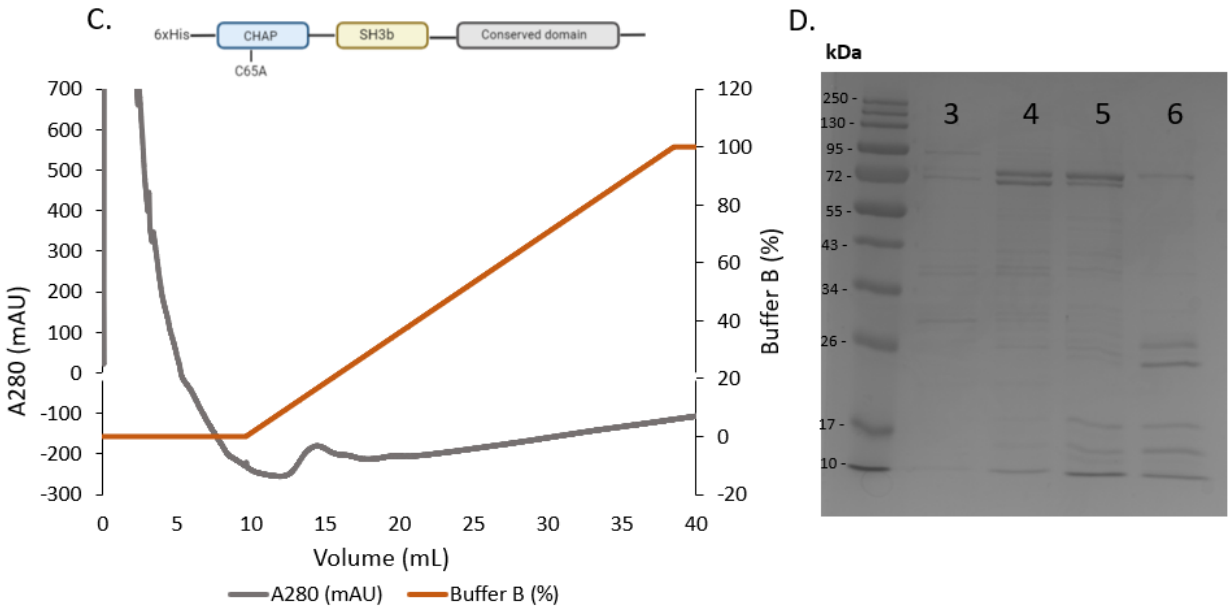

Figure A.6.2: C. Purification of CbpD^{C65A} with the graph displaying 280 nm absorbance of the protein (grey line) after about 3 mL of linear gradient elution with Buffer B (500 mM free imidazole) (orange line). Fractions 3-6 were therefore collected. **B.** SDS-PAGE does not confirm the presence of CbpD^{C65A} as the correct band sizes is not displayed (40.01 kDa).

Appendix 7: Complete growth curve with CbpD^{G302A,G303A}

Figures A.7.1 and A.7.2 display the growth curves of *Streptococcus dysgalactiae* strain MA201 with the addition of CbpD^{G302A,G303A} at OD_{550} = 0.2. Iodoacetamide-treated CbpD^{SD} was used as negative control and a CbpD^{SD}-wild-type was used as positive control. Different concentrations of the various CbpD proteins were used, all of which are included in the figures below. Untreated cells were also included. Figure A.7.1 displays the OD_{550}, while Figure A.7.2 displays the Sytox™ Green fluorescence. The 1 µg/mL concentration of CbpD was seen to give the ideal lysis.

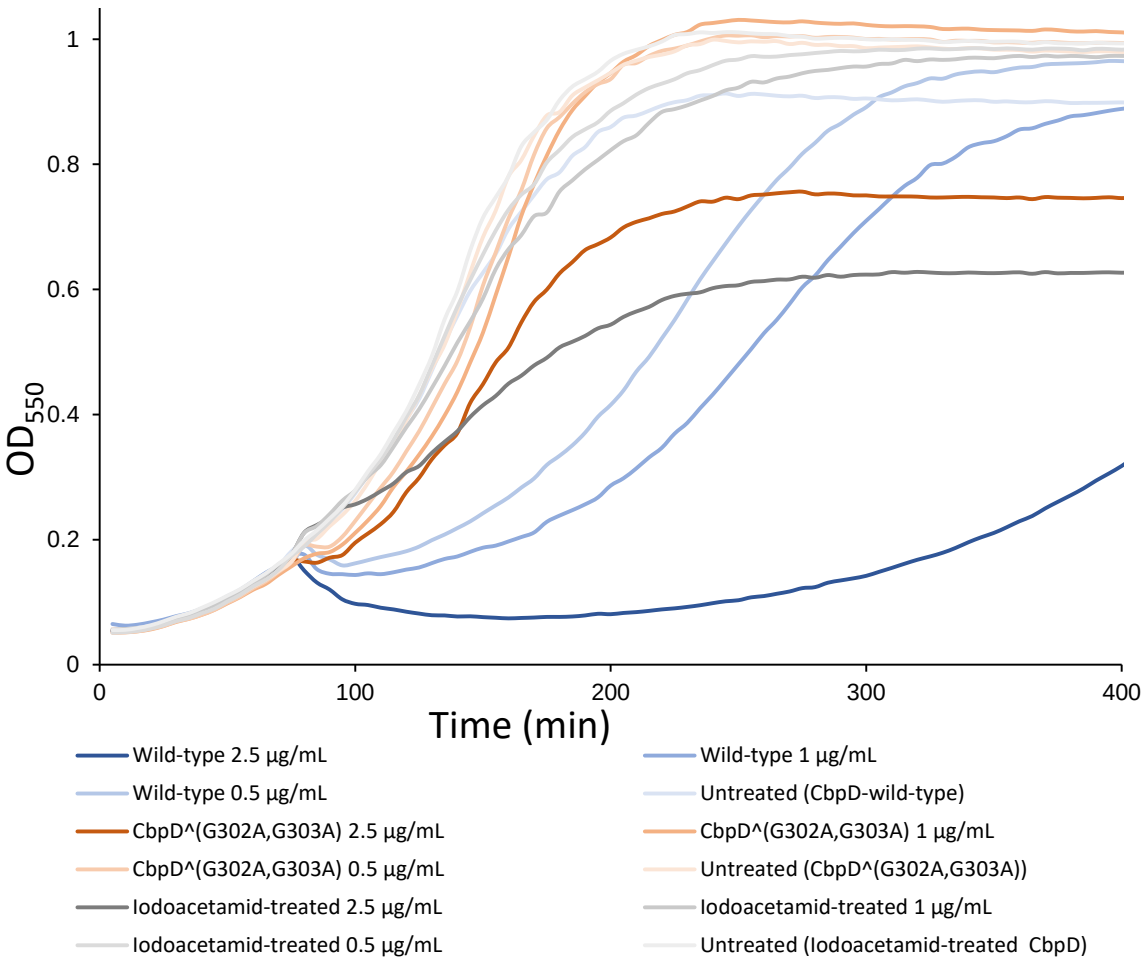

Figure A.7.1 The general trends is that adding iodoacetamide-treated CbpD^{SD} and CbpD^{G302A,G303A} at all concentrations at OD_{550} = 0.2 cause no significant lysis of the *Streptococcus dysgalactiae* MA201 cells, while adding CbpD^{SD}-wild-type cause significant lysis. The LAGG motif therefore seems important for guiding the CbpD^{SD} towards cell septum, and that this localization is necessary for cell rupture.

Appendix 7: Complete growth curve with CbpDG302A,G303A

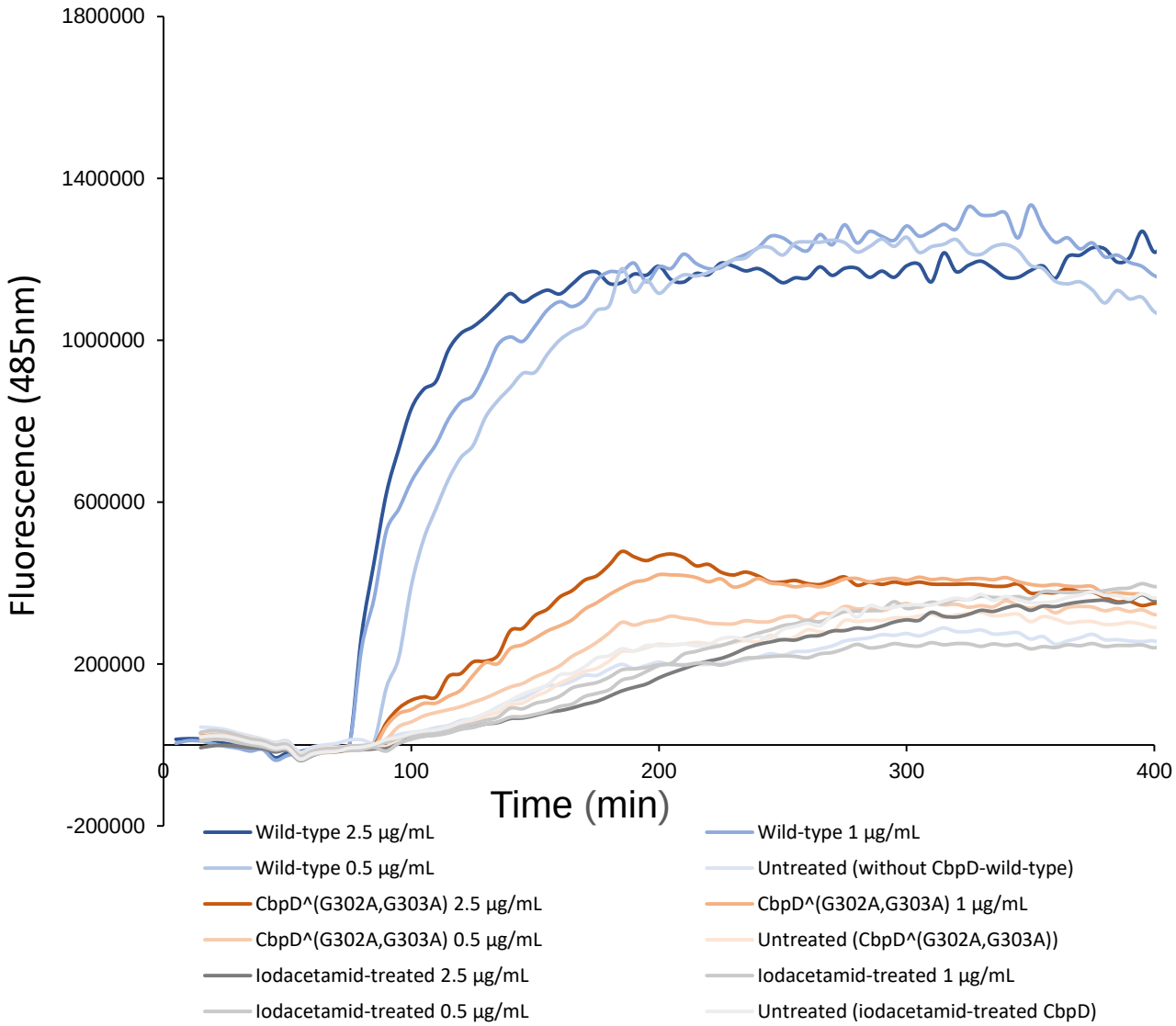

Figure A.7.2: The general trends is that adding iodoacetamide-treated $CbpD^{SD}$ and $CbpD^{G302A,G303A}$ at all concentrations at OD_{550} = 0.2 cause no significant lysis of the *Streptococcus dysgalactiae* MA201 cells as indicated by the low fluorescence signals, while adding $CbpD^{SD}$-wild-type cause significant lysis as shown by higher fluorescence due to more lysis.

Appendix 8: Immunity test of *Streptococcus dysgalactiae* strain 047

To check whether *S. dysgalactiae* produced immunity against CbpD^{SD}, strains 015 and 047 was used. The summarized results for strain 015 can be found in section 4.5, whereas results from strain 047 is presented here. The resulting growth curve showed not much difference between the induced cells and the non-induced cells after adding ComS2 (at OD_{550} = 0.2) and CbpD^{SD} (after waiting 20 minutes). In addition, it was observed that the strain did not grow as smoothly as strain 015.

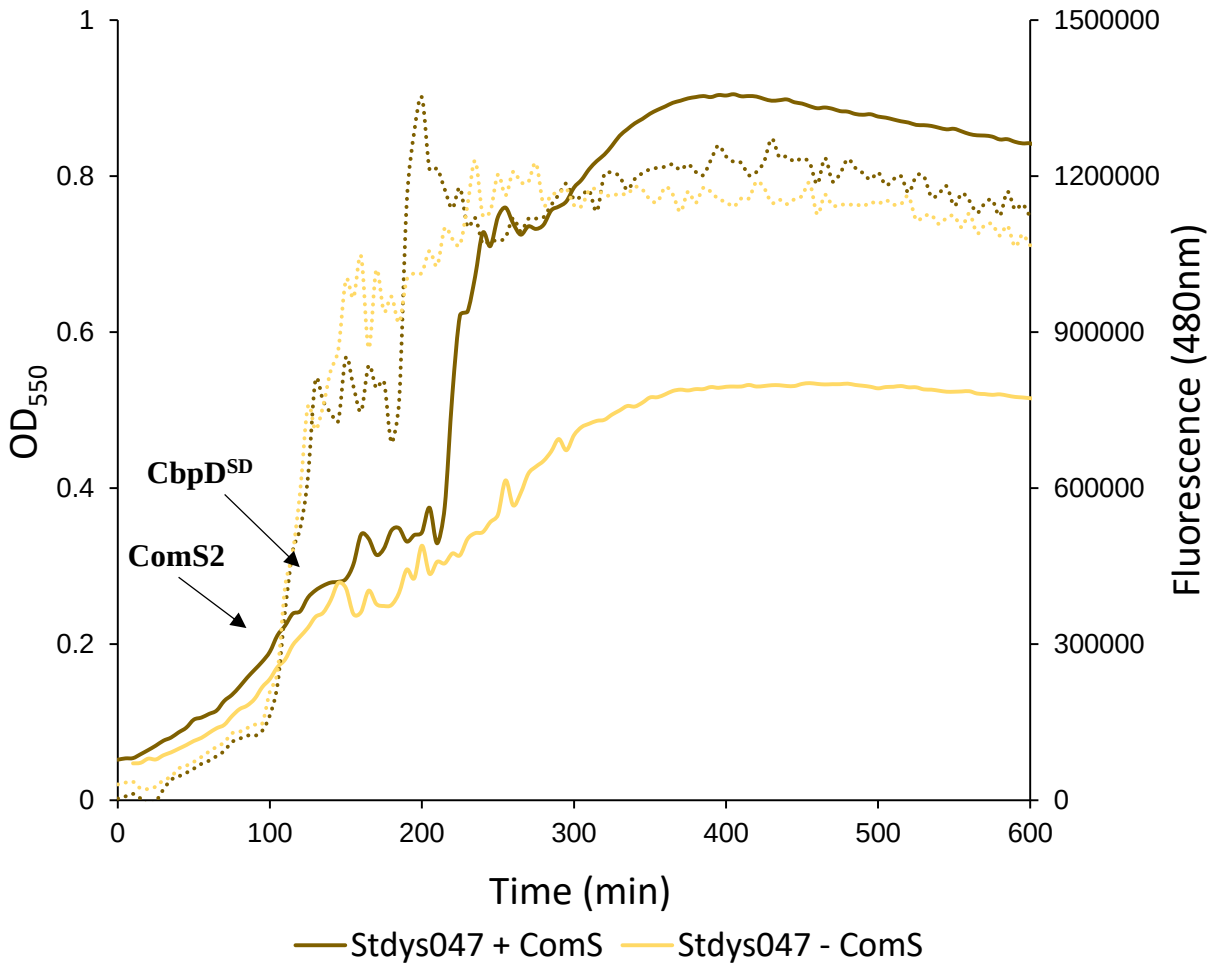

Figure A.8: Growth curve assay *Streptococcus dysgalactiae* strain 047 cells where ComS2 (250 ng/mL) (half of wells) was added at OD_{550} = 0.2, and CbpD^{SD}-wildtype (1 µg/mL) added after 20 minutes. The graph displays a dark yellow line (OD_{550}) and stapled line (fluorescence) representing the induced 047 cells, while the non-induced cells are the light yellow lines. There is no clear difference between the induced and non-induced cells for either OD_{550} measurements or fluorescence, although a sharp increase in OD is seen for the induced cells.

Appendix 9: Multiple sequence alignment of CbpDs from strains 015, 047 and MA201

A multiple sequence alignment was produced from ClustalW (165) between CbpD sequences from *S. dysgalactiae* strains MA201, 015 and 047. It is seen that the CbpDs are highly homologous as almost all amino acids between the strains match, and only a few areas where some of the amino acids are only similar or do not match. It was therefore believed that CbpD^{SD} from *S. dysgalactiae* strain MA201 would work on *S. dysgalactiae* strains 015 and 047.

```
MA201         MKKIHQLLVSGAILLSVNGAVSSVASTLNAEHTGVVHAAVLGDNYPSKWKKGSGIDSWNM      60
Stdys015      MKKIHQLLVSGAILLSVNGAVSSVASTLNAEHTGVVHAAVLGDNYPSKWKKGSGIDSWNM      60
Stdys047      MKKIHQLLVSGAILLSVNGAVSSVASTLNAEHTGVVHAAVLGDNYPSKWKKGSGIDSWNM      60
              ************************************************************

MA201         YVRQCTSFVAFRLSSANGFQLPKGYGNACTWGHIAKKQGYTVNKTPKVGAVAWFDTNAFQ      120
Stdys015      YVRQCTSFVAFRLSSANGFQLPKGYGNACTWGHIAKKQGYTVNKTPKVGAVAWFDTNAFQ      120
Stdys047      YVRQCTSFVAFRLSSANGFQLPKGYGNACTWGHIAKKQGYTVNKTPKVGAVAWFDTNAFQ      120
              ************************************************************

MA201         SHATYGHVAWVAEVRGDSVVIEEYNYNAGQGPEKYHKRQIPKNHVSGYIHFKDLPSGEAS      180
Stdys015      SHATYGHVAWVAEVRGDSVVIEEYNYNAGQGPEKYHKRQIPKNHVSGYIHFKDMLSGEAS      180
Stdys047      SHATYGHVAWVAEVRGDSVVIEEYNYNAGQGPEKYHKRQIPKNHVSGYIHFKDMLSGEAS      180
              ************************************************: *****

MA201         KSQTKEQQVSKEAVKQGGTYHFTERTPVKAQAQLTSPDLAYYNPGQSVHYDQAMTVDGHE      240
Stdys015      KSQTKEQQVSKEAVKQGGTYHFTERTPVKAQAQLTSPDLAYYNPGQSVHYDQAMTVDGHE      240
Stdys047      KSQTKEQQVSKEAVKQGGTYHFTERTPVKAQAQLTSPDLAYYNPGQSVHYDQAMTVDGHE      240
              ************************************************************

MA201         WISYLSFSGNRRYIPIKKIEQKTQQVSETTSPINIGDRVTFPGVFRVDRIVNNLLVSEEL      300
Stdys015      WISYLSFSGSRRYIPIKKTEQKTQQVSETTSPINIGDRVTFPGIFRVDRIVNNLLVSEEL      300
Stdys047      WISYLSFSGSRRYIPIKKTEQKTQQVSETTSPINIGDRVTFPGVFRVDRIVNNLLVSEEL      300
              *********.******** ************************:****************

MA201         AAAGATSLNWIDPSPLDETDRKGVKAGNQILQAGEFFVIPGNYRVLKVDRPSNGIYVKIG      360
Stdys015      AGGGATSLNWIDPSPLDETDRKGVKAGNQILQAGEFFVIPGNYRVLKVDRPSNGIYVKIG      360
Stdys047      AGGGATSLNWIDPSPLDETDRKGVKAGNQILQAGEFFVIPGNYRVLKVDRPSNGIYVKIG      360
              *..************************************************************

MA201         SRGTWLTADKASKLP   375
Stdys015      SRGTWLTADKASKLQ   375
Stdys047      SRGTWLTADKASKLQ   375
              **************
```

Figure A.9: Multiple sequence alignment between CbpD sequences obtained from *S. dysgalactiae* strains MA201, 015 and 047. Their CbpDs are almost identical with each other due fully conserved amino acids at almost all positions, with only few areas of mismatches.

Norges miljø- og biovitenskapelige universitet
Noregs miljø- og biovitskapelege universitet
Norwegian University of Life Sciences

Postboks 5003
NO-1432 Ås
Norway